

Standards for papers on cloning

In the wake of the Hwang scandal, journals have been reviewing their refereeing procedures. Following a survey of experts, here are *Nature's* thoughts on papers about cloning, with an invitation to comment.

Cloning cells or organisms by nuclear transfer is laborious, time-consuming and expensive. The timescale for replication of published results may be on the order of years. But conversely, a successful cloning experiment has the advantage, rare in science, that it can usually be evaluated relatively easily, by DNA testing of tissues collected during the procedure.

Peer review aims to assess credibility but is by and large incapable of detecting dishonesty (although *Nature* is currently reviewing its procedures with respect to image manipulation). With the caveat that peer review will continue to be based largely on trust, what data can be regarded as sufficient proof of cloning? Before the Hwang scandal, the gold standard for proving that an animal had been cloned from nuclear transfer was to test for identical nuclear DNA fingerprints of the clone and nuclear donor. On the other hand, mitochondrial DNA from the cloned animal should differ from that of the nuclear donor, providing a straightforward way of ruling out sample mishandling or outright fraud.

In the light of all that has happened, we think it sensible from now on to ask authors to provide not only nuclear but also mitochondrial DNA fingerprints for all cloning papers submitted to *Nature*. It should be noted, however, that there may be confounding factors in interpreting such data. For instance, the mitochondrial contribution of the nuclear donor may vary depending on the species, on whether it is an interspecies hybrid, and on the nuclear-transfer technique used. In the case of papers reporting new embryonic stem-cell lines, nuclear DNA fingerprints of the lines should be presented for comparison with existing lines, to help rule out sample mishandling (intentional or accidental) or contamination with other cell lines.

How much data should be provided when papers are submitted? Authors of cloning papers should always present enough data to document the logical flow and efficiency of a cloning procedure. But we may in addition require authors to provide raw data on request, for inspection by reviewers or editors. This allows an additional level of verification should questions arise during the review process, by ensuring that the data presented in the paper are an accurate interpretation of the raw data.

Independent tests

In the light of the extraordinary circumstances surrounding the Hwang case, *Nature* commissioned an independent scientist to verify that the dog Snuppy (B. C. Lee *et al.* *Nature* **436**, 641; 2005) is indeed a clone, by DNA fingerprinting analysis of blood samples from Snuppy and the nuclear donor. The indications are positive and the results are being peer-reviewed as we go to press. However, a few scientists have suggested that *Nature* should make such independent tests a condition of publishing cloning papers.

After weighing this suggestion carefully, we concluded that imposing such a standard on the cloning field as a condition of publication

would be an overreaction, and one with myriad inherent logistical problems. For example, who will pay for testing, how will mishandling of samples be prevented, and how will the scientists running the verification tests be acknowledged for what could be significant additional work? Moreover, because the cloning field is not unique in its susceptibility to fraud, it would follow that we should require independent verification of all our papers, which is untenable in the current system. The gains of detecting the rare cases of fraud would be negated by the impediments to publication this would bring.

However, in the best interests of science, we encourage researchers embarking on landmark cloning studies to seek independent verification themselves, and to include a report of these findings in their initial submission. And keeping in mind the principle that extraordinary claims require extraordinary proof, *Nature* may in rare cases demand it.

Nature and most other journals require as a condition of publication that authors make relevant reagents available to the scientific community, and encourage the deposition of cell lines and mutants in established repositories. But it has now been suggested that journals take this a step further, and require the deposition of critical samples in repositories such as the American Type Culture Collection, where they can be cheaply and rapidly made available to other researchers following publication. We are currently weighing up the practicalities of implementing such a requirement, and we would welcome feedback on whether we should and, if so, how to best do it.

Regardless of what repository is used, we urge scientists embarking on what are likely to be landmark cloning studies to ensure that critical samples are properly stored for later verification. As part of this procedure, an independent scientist not involved in the study should obtain and store cells from the nuclear donor, oocyte donor and the resulting animal or stem-cell line (or oversee their deposition in a repository). This precaution is especially important in the case of human donors, where it may not be possible to go back to the subjects to obtain additional tissues for later verification. Funding agencies should make granting dependent on procedures to ensure later verification of samples, and institutions should demand this for approval by the institutional review board.

The Hwang debacle reminds us that science is largely a self-correcting process in which scientists, editors, reviewers, journalists, funding agencies and institutions all play crucial corrective roles. In the aftermath of this deception, we should all undertake close scrutiny of our procedures and standards, with an eye towards preventing it from ever happening again. We should be vigilant against knee-jerk reactions and witch hunts. *Nature* welcomes feedback on its approach (e-mails may be sent to authors@nature.com).

"Keeping in mind that extraordinary claims require extraordinary proof, *Nature* may in rare cases demand it."



A new ERA?

A novel component of the European Research Area will require national funders' cooperation.

Anyone who knows what COST, EURYI, ESF, ELSON, ESOF, EUROHORCS and FENS are probably spends their days deciphering European research policy. These bodies — and there are many more — represent independent Europe-wide initiatives aimed at encouraging scientists and research funding agencies to think European rather than national.

When former European research commissioner Philippe Busquin coined the term European Research Area (ERA) in 2000, he was endorsing such efforts and committing the European Commission to the common aim. But he was also thinking bigger. The vast majority of research money in Europe is in the hands of national funding agencies, which mostly do not allow it to be spent in other countries. In an ideal ERA, national agencies would see the value of sharing much more of their funds in activities for which a larger European scientific community makes sense.

The persistent resistance of many countries in the European Union to this ideal is unsurprising. Nevertheless, the commission has just launched another step towards it: Eurobiofund. This is a new forum that will bring together public and private research funding bodies to listen to pitches from European bioscientists.

The scientists will present hot areas of basic research that they believe need trans-national support if Europe is to remain competitive with the United States and Asia. Funding bodies could sign up to a specific theme, such as lipidomics, and put out a joint call for proposals to be handled by a common evaluation system. They will not necessarily create a shared pot of money, but each could fund

their own scientists who win the open competitions. As an incentive for national agencies to flout their own traditions in this way, the commission will top up any joint project with money from its Seventh Framework Programme of Research, which is to be launched at the end of this year.

Funding organizations have signed up to the principle of the Eurobiofund, with the first forum to take place this November in Finland. The commission has given the European Science Foundation (ESF) €1 million (US\$1.2 million) to set up and run the initiative.

The success of this experiment depends on many factors — in particular, whether its budget is confirmed within the Framework programme, whose detailed contents will be defined later this year.

But it also depends on whether agencies are genuinely ready to support joint evaluation procedures. The experience with EURYI — the European Young Investigator Awards, which are also administered by the ESF and established by EUROHORCS (the European Heads of Research Councils) — provides grounds for only cautious optimism. Against historical odds, Germany's research council, the DFG, managed to persuade its government to pay into a common financial pot, but then found that other agencies had failed to get similar agreements. As a result, EURYI winners must be funded by their own national agencies. And to make matters worse, British funding agencies have already pulled out of the scheme.

Despite this, Eurobiofund is a positive sign of the commission's willingness to generate ideas for the European Research Area and serve as a catalyst. It may end up being just a small step towards the ideal, but it is the biggest single step that we have seen for some time. European scientists should give it their full support. ■

"The European Commission will top up any joint project with money from the seventh Framework programme."

Circulation challenge

The lack of monitoring of ocean currents must be addressed quickly.

The idea of a 'mini ice age' triggered by a shutdown of the oceans' thermohaline circulation (THC) has been rich fodder for dramatic scenes from Hollywood to the Pentagon. The currents of the THC take cold surface water from high latitudes southwards at depth, driving low-latitude warm surface waters north. This powerful heat conveyor belt is driven by differences in salt-water density. According to models, if sufficient fresh water is added to the ocean, which could happen as a result of global warming, the THC may cease. Indeed, palaeoclimate studies suggest that the THC has shut down a number of times during colder climates in the past 100,000 years.

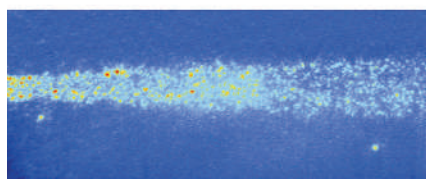
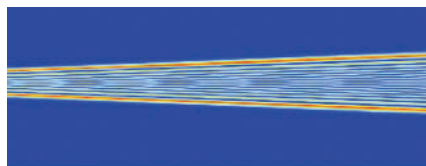
But most climate researchers have long abandoned the notion of isolated cold regions amidst a globally warming world. It now seems less likely that even a full collapse of the THC — which, although improbable, might still occur towards the end of the century — would significantly cool Britain or Scandinavia (see page 256).

The matter is not yet closed, however. A weakening of the THC — and recent observations published in *Nature* have suggested that the currents have begun to change — may lead to perturbations in global climate systems, with unknown side effects. There are many uncertainties, but it is clear that people in Western Europe and eastern North America are less threatened by a consequent rapid climate change (and are more capable of adapting to it) than many people in poorer societies.

More measurements are clearly needed if we are to fill the enormous gaps in our knowledge of ocean behaviour. Autonomous observation tools, such as drifting floats and moored buoys, are now allowing scientists for the first time to monitor the state of the ocean currents almost in real time. This is an important advance, but observations must be sustained for much longer periods than foreseen in the six-year RAPID programme (see www.noc.soton.ac.uk/rapid/rapid.php). Furthermore, they should be augmented globally if we are to anticipate possible changes in ocean behaviour with any confidence. ■

"A weakening of the thermohaline circulation may lead to perturbations in global climate systems, with unknown side effects."

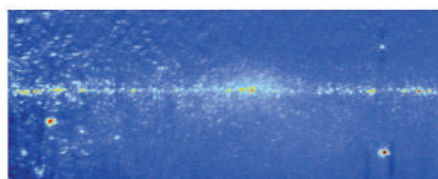
RESEARCH HIGHLIGHTS



Ray of light

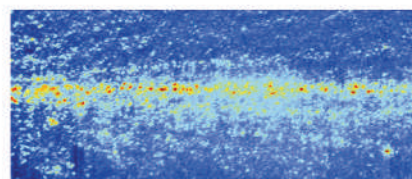
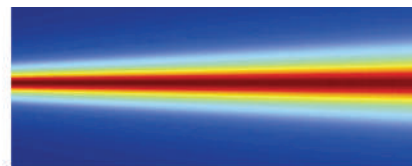
Nature Mater. 10.1038/nmat1568 (2006)

The spreading of a beam of light as it propagates has made using optics to shuttle information across silicon chips problematic. But now a team at the Massachusetts Institute



of Technology in Cambridge has stopped diffraction over a distance of one centimetre.

Researchers have previously used photonic crystals, which are drilled with holes that influence their optical properties, to control diffraction over very short distances. By optimizing the arrangement of holes in a



silicon chip, Peter Rakich, Marcus Dahlem and their colleagues created a straight, or 'supercollimated', beam 100 times longer than achieved in the past.

The experimental results (bottom) matched theoretical predictions (top) for how different wavelengths of light travel through the crystal.

NATURE MATER.

CHEMISTRY

Weird water

J. Phys. Chem. B doi:10.1021/jp056198x (2006)

In work that bravely ventures into the realms of pseudoscience, researchers have shown that magnetic fields can affect the properties of water containing dissolved oxygen.

Ichiro Otsuka and Sumio Ozeki of Shinshu University in Nagano, Japan, performed a series of experiments on distilled water in a vacuum and on water exposed to oxygen, before and after the application of magnetic fields. The magnetic treatment did not change the properties of distilled water but it did change the behaviour of the oxygenated water, including introducing a new feature to its absorption spectrum. The results do not, of course, endorse the weird and wonderful claims made by purveyors of magnetized water, but they do raise questions about how magnetic fields affect water's structure.

NEUROBIOLOGY

Forgetful flies

Cell **124**, 191–205 (2006)

Researchers have homed in on a molecular pathway in the junctions between nerve cells, known as synapses, that regulates the formation of long-term memory.

Sam Kunes of Harvard University in Cambridge, Massachusetts, and his colleagues targeted a protein complex called RISC. This forms part of the RNA interference mechanism, helping short RNAs known as microRNAs to switch off genes.

Working in the fruitfly *Drosophila*, they showed that one of the RISC proteins, Armitage, has to be destroyed at particular synapses for the protein synthesis that underpins memory to occur. Furthermore, Armitage-mutant flies had impaired long-term memory. The RISC pathway is also found in mammals, so the same mechanism may operate in humans.

colleagues at the University of Zürich, Switzerland, studied Australian desert ants (*Melophorus bagoti*, pictured). Using artificial barriers, they forced the ants to learn different routes to and from a food source. Homeward-bound ants that were transferred to a point along their outward route could not follow it in reverse, but instead began searching behaviour.

CELL BIOLOGY

Centrefold proteins

Cell **124**, 75–88 (2006)

Efficient machinery for protein origami may have allowed eukaryotes to evolve more complex proteins than prokaryotes such as bacteria.

Prokaryotes are thought to use a single set of molecular chaperones to help fold up both newly synthesized proteins and ones that have lost their shape due to heat or other stress. Judith Frydman of Stanford University in California and her team now show that eukaryotic cells have distinct networks for each of these tasks.

They found that in eukaryotic cells some chaperones are dedicated to folding newly made proteins. This may have facilitated the evolution of multidomain proteins.

ASTRONOMY

Phantom galaxy

Astrophys. J. **636**, 575–581 (2006)

A galaxy thought to be the most distant ever observed was a mirage, according to results from the orbiting Spitzer Space Telescope.

ANIMAL BEHAVIOUR

One-way street

Curr. Biol. **16**, 75–79 (2006)

Foraging insects such as bees and ants have a remarkably developed sense of where they are going and how to get home again. But new research shows that for ants, at least, this ability does not extend to knitting knowledge of routes into maps: ants rely on memorized routes that only work in one direction.

Rüdiger Wehner and his



Abell 1835 was identified in data from the Very Large Telescope in Chile, using a technique that relied on its light being gravitationally magnified by a massive galaxy cluster. It appeared to have formed just 500 million years after the Big Bang, at a redshift of ten.

The Spitzer observations from Graham Smith of the California Institute of Technology, Pasadena, and his colleagues follow other non-detections of Abell 1835, sealing the case against the galaxy. However, Smith's team says this failure does not invalidate the lensing technique.

IMMUNOLOGY

Irresistible infection

Nature Immunol. doi:10.1038/ni1300 (2006)

The puzzling observation that people with malaria or systemic blood infections have suppressed immune systems may be explained by new data.

A team led by researchers at the Walter and Eliza Hall Institute of Medical Research in Parkville, Australia, traces the effect to dendritic cells — the sentinels of the immune system (pictured above, surrounded by malaria-infected blood cells).

Dendritic cells capture and present viral antigens to prime other immune cells. But this process, known as cross-presentation, was inhibited in mice infected with the malaria parasite. The work links immunosuppression to over-stimulation of the dendritic cells, via their Toll-like receptors, causing the cells to mature to a stage at which they can no longer cross-present.

SYNTHETIC BIOLOGY

Absolute minimum

Proc. Natl Acad. Sci. USA **103**, 425–430 (2006)

Researchers have revised their estimate of the minimal set of genes needed to sustain bacterial life. The findings will guide scientists hoping to build a cell from scratch.

In 1999, a team led by J. Craig Venter announced that they had mutated every gene in the bacterium *Mycoplasma genitalium* — which has the smallest genome of any free-living organism — and estimated that it needs between 265 and 350 of its 517 genes to grow (*Science* **286**, 2165–2169; 1999).

But their estimate was approximate because they did not isolate and study each mutant strain individually. Now, follow-up research led by Hamilton Smith at the J. Craig Venter Institute in Rockville reveals that the minimum genome consists of 387 protein-coding and 43 RNA-coding genes.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

D. FERGUSON, ISM/SPL

NETWORKS

Exclusive clubs exposed

Nature Phys. doi:10.1038/nphys209 (2006)

"The rich are different from you and me," said F. Scott Fitzgerald. But do the rich — the well connected and highly influential — really form an exclusive club?

In network theory, 'rich' nodes are the most highly connected, and the 'rich-club' phenomenon, where rich nodes are preferentially connected to one another, has been proposed to exist in many systems.

Alessandro Vespignani and his colleagues at Indiana University in Bloomington report a new approach to measuring rich-club ordering. They find that rich clubs genuinely exist among scientists, but for the Internet and protein networks there is actually less communication among rich nodes than expected by chance.

IMMUNOLOGY

Antibodies run amok

Proc. Natl. Acad. Sci. USA **103**, 281–286 (2006)

The cellular damage that underlies multiple sclerosis could be caused by an 'abzyme' — an antibody that acts as an enzyme — say researchers.

Multiple sclerosis is characterized by degradation of the myelin sheath, an insulating layer of protein that helps brain cells to transmit electrical signals. Alexander Gabibov of the Shemyakin and Ovchinnikov Institute of Bioorganic Chemistry in Moscow, Russia, and his colleagues ran *in vitro* tests of antibodies to myelin, isolated from the blood of multiple sclerosis sufferers, and found that the antibodies are able to cleave myelin protein.

The authors suggest that inhibiting the antibodies' activity could form a previously unrecognized approach for treating multiple sclerosis.

JOURNAL CLUB

John P. Moore

Weill Medical College of Cornell University, New York, USA

An HIV researcher finds hope in a study of male circumcision.

Many of us working on HIV have acquired 'long-term non-progressor' status over the past 10–20 years. We have failed to make an effective vaccine, we don't yet have a microbicide, and global access to antiretroviral drugs is still problematic. So it's good to read of anything that could help to reduce the global spread of this virus.

I was therefore intrigued by a report of a trial conducted in South Africa, involving 3,274 men, which showed that circumcising males significantly reduced their susceptibility to HIV-1 infection (B. Auvert *et al.* *PLoS Med.* **2**, e298; 2005).

The underlying mechanism remains to be determined, although I think it is likely to have a biological basis that is related to the presence in the foreskin of target cells that are particularly susceptible to HIV infection.

Could there be a behavioural explanation? After hearing a presentation of this work, a Jewish friend commented that following his circumcision, he desisted from risky sexual practices for about 20 years. But in the trial, sexual activity actually increased slightly in the circumcised group.

Seriously, though, any practical intervention that could reduce HIV transmission is to be welcomed. And I was encouraged by the editorial and commentary that accompanied the paper. These articles discuss the background to the study and counter any criticisms that it has ethical flaws.

Sometimes Western ethicists feel the need to pontificate on how studies should be carried out in Africa. Their ivory-tower ethics can be very unhelpful given the scale of the AIDS epidemic nowadays.

Local scientists and review boards are capable of deciding for themselves what is right and wrong.

NEWS

Alarms ring over bird flu mutations

Scientists studying virus samples from the human outbreak of avian flu in Turkey have identified three mutations in the virus's sequence. They say that at least two of these look likely to make the virus better adapted to humans.

The Turkey outbreak is unusual, because of the large family clusters of cases; the fact that many of those infected have only mild symptoms; and the speed with which infections have arisen — twenty cases, including four deaths, in less than two weeks. So scientists are urgently trying to establish whether the virus is behaving differently in this outbreak from previous ones in Asia. In particular, international teams are investigating the possibility that the virus is moving between people.

"With such a large number of cases within such a short period of time, human-to-human transmission is something that we've had to consider," says Maria Cheng, a spokeswoman at World Health Organization (WHO) headquarters in Geneva.

As *Nature* went to press, samples from the first two teenagers in the country to die had been sequenced by a WHO collaborating centre at the National Institute of Medical Research (NIMR) in London.

The results so far are not comforting. The first mutation found, announced last week, involves a substitution in one sample of an amino acid at position 223 of the haemoagglutinin receptor protein. This protein allows the flu virus to bind to the receptors on the surface of its host's cells.

This mutation has been observed twice before — in a father and son in Hong Kong in 2003, and in one fatal case in Vietnam last year. It increases the virus's ability to bind to human receptors, and decreases its affinity for poultry receptors, making strains with this mutation better adapted to infecting humans.

The same sample also contained a mutation at position 153 of the haemoagglutinin protein, *Nature* has learned. Cheng says this information was not included in WHO statements, because "it is not clear what role this particular change plays".

Finally, both samples from the Turkish teenagers show a substitution of glutamic acid



The recent outbreak of bird flu in Turkey has thrown up viruses with mutations that threaten humans.

with lycine, at position 627 of the polymerase protein, which the virus uses to replicate its genetic material. This mutation has been seen in other flu sequences from Eurasian poultry over the past year. It was also present in the one person who died during an outbreak of H7N7 in the Netherlands in 2003, and in a few

"Human-to-human transmission is something that we have had to consider."

people in Vietnam and Thailand.

The polymerase mutation is one of the ten genetic changes that gave rise to the 1918 pandemic flu virus. Like the 223-haemoagglutinin mutation, it signals adaptation to humans, says Alan Hay, director of a WHO influenza laboratory at the NIMR. "There is this glutamic acid-lysine

O. ORSAL/AP

Doctor admits *Lancet* study is fiction

A Norwegian researcher dreamed up the lives and lifestyles of some 900 people — and used them in a study on cancer. Then, last October, Jon Sudbø had his results published in *The Lancet*.

The revelation comes hard on the heels of the Woo Suk Hwang scandal, in which several important advances in human cloning reported by the South Korean researcher turned out to be faked.

The blatant nature of Sudbø's fiction emphasizes questions already being asked about the effectiveness of peer review, even in top journals, and about who should be responsible for catching fraud (see page 243).

The latest deception was discovered by

Camilla Stoltenberg, a director of epidemiology at the Norwegian Institute of Public Health in Oslo, who was catching up on her literature reading over Christmas. Sudbø's paper claims to analyse a public-health database and show that taking anti-inflammatory drugs can reduce the incidence of mouth cancer (J. Sudbø *et al.* *Lancet* 366, 1359–1366; 2005). "I was surprised because it refers to the Cohort of Norway, for which I am responsible," Stoltenberg says. She knew that this could not have been the source of the lifestyle data as the paper claimed.

Last week, Sudbø, who is based at the Norwegian Radium Hospital in Oslo,



**WEB USERS JUDGE SITES
IN THE BLINK OF AN EYE**
Potential readers make
snap decisions in just
50 milliseconds
www.nature.com/news

flip," he explains. "Glutamic acid is associated with flu-virus replication in birds, and lysine is in primates."

The Turkey strains are the first in which the polymerase and receptor-binding mutations have been found together. They could make it easier for humans to catch the virus from poultry. But they might also favour human-to-human transmission. This is because the polymerase change helps the virus to survive in the cooler nasal regions of the respiratory tract, and the haemoagglutinin mutation encourages the virus to target receptors in the nose and throat, rather than lower down in the lungs. The virus is thought to be more likely to spread through droplets coughed from the nose and throat than from infections lower down.

Hay points out, however, that it is difficult to predict how the mutations will actually influence the virus's behaviour. He adds that just two changes are unlikely to create efficient human-to-human transmission on their own.

Establishing what effects these changes are having on the epidemiology of the current outbreak is a top priority for research teams working in Turkey. "We must learn more about the mild cases and be absolutely sure of whether these viruses are behaving differently from those we have seen elsewhere," says Hay. "It is early days in terms of what we know about the viruses causing these infections."

Researchers are sequencing more strains from the Turkey cases, to see whether they share the mutations and to check for further changes. Samples were expected to arrive in London on 18 January, after being held up for more than a week in Turkey because of the Eid ul-Adha holiday period. ■

Declan Butler

Will Germany choose a fair elite?

MUNICH

Germany's universities are taking part in a contest that is intended to boost their global standing. But claims that political bias will affect the selection process have sparked debate over whether the top institutions will — or even should — win.

Despite being one of the world's largest economies, Germany's research universities punch below their weight in world rankings. So, in 2004, the then centre-left government created a €1.9-billion (US\$2.3-billion) scheme to name and reward a handful of elite universities.

Each chosen university will receive up to €30 million a year for five years to help them compete with the likes of Harvard, Cambridge and Tokyo. Winners of the first round will be announced on 20 January.

But last week brought allegations that the contest will not be based purely on academic merit. Daniel Guhr, founder of the Illuminate Consulting Group, based in San Diego, California, had

been helping Berlin's Humboldt University to pep up its application. In an interview with Berlin newspaper *Der Tagesspiegel*, he claimed that the competition is bound to be biased by political and regional considerations, and he released details of a study in

"I'm just saying what everyone already knows." which he predicts the universities most likely to win.

For example, Guhr argues that universities with historically strong reputations will have an advantage regardless of their current research status. And that institutes such as the Free University of Berlin and the University of Würzburg in Bavaria — both with strong research programmes — do not have a fair chance because they are geographically close to the favourites, including Humboldt and the Technical University of Munich. "If the decision were solely about scientific excellence, a disproportionate share of the winners would be in southern Germany," he told *Nature*. "Clearly, this would be hard to accept in Germany's federal system." ■

Guhr adds that he sees such factors as "legitimate and reasonable" — without a political component to the decision, even the strongest universities in the east, for example, the University of Dresden, would have little chance. But he says that biases should be properly acknowledged because of the consequences for the losers. "Funding and attention will focus on the new elite."

The consultant has been almost unanimously criticized for his outburst. Humboldt cut ties with him, and research managers throughout Germany have rejected the idea. "The claims are totally unfounded," insists Eva-Maria Streier, spokeswoman for Germany's main research agency, the DFG, which is running the competition. "Political interests will not play a role in the outcome."

Guhr says he is surprised by the response. "I'm just saying what everyone already knows," he says. Many will be keeping an eye on his predictions when the final results are announced. ■
Robert Rentzsch

NORWEGIAN RADIUM HOSPITAL



Novel approach:
Jon Sudbø
simply invented
his test subjects.

admitted that the data had not come from that database or any other, but from thin air.

Many details of this case still need to be worked out. There is some indication that Sudbø may have mental health problems. It

is also not clear what his 13 co-authors knew about the fraud — the paper identifies three others as contributing equally to the research, and among the other co-authors are Sudbø's wife and his identical twin. None of the authors could be reached for comment.

The hospital has asked that Sudbø's other work be examined in an independent investigation, to be set up this week by Anders Ekbom, an epidemiologist at the Karolinska Institute in Stockholm, Sweden. But the case is already set to change research policy in Norway. The country's health minister, Sylvia Brustad, announced on 16 January that previously stalled reforms on medical research will probably be law by

the autumn. The new rules would put more responsibility for catching fraud on the shoulders of the institutions where the research was done.

Richard Horton, editor of *The Lancet*, insists his journal is not at fault. "This is all so similar to the Hwang thing that we have just been through," he says wearily. "Peer review is a great system for detecting badly done research, but if you have an investigator determined to fabricate an entire study, it is not possible to pick it up." The mechanism of peer review at his journal is currently being examined as part of the largest study ever conducted into the process (see page 252). ■
Emma Marris

Astronomers told to cut it out

WASHINGTON DC

There's no point in complaining: some US ground-based telescopes almost certainly must close if the country's astronomy spending is to be brought under control.

That was the bleak message delivered to a room of anxious astronomers on 10 January by officials of the National Science Foundation (NSF) during the American Astronomical Society's annual meeting in Washington DC. Faced with flat budgets and growing expenditures, the foundation has asked a panel of 13 senior scientists to find \$30 million in savings out of the astronomy division's \$195-million annual budget — by any means necessary.

"The problem is huge," Wayne Van Citters, director of the foundation's astronomy division, told the assembled scientists. "We're rapidly outstripping our ability to operate the things that we are building."

The agency is the main government supporter of ground-based astronomy in the United States. It operates several major facilities,

including the Gemini, Kitt Peak, Arecibo and National Solar observatories, and the National Astronomy and Ionosphere Center.

In addition to these telescopes, the NSF is planning several costly, next-generation projects. It is putting nearly \$50 million a year towards constructing the Atacama Large Millimeter Array (ALMA), an international radio telescope set to be completed in 2012. And it recently committed \$14 million to developing the Large Synoptic Survey Telescope, a project to digitally record much of the night sky.

Money is allocated separately for the construction of these and other facilities, but Van Citters says it is unclear how the agency will operate them once they are built. In the fiscal year 2006, the foundation's entire maths and physical sciences budget grew by only \$16 million, about half of the amount it will need annually just to operate ALMA.

By slicing \$30 million from existing facilities, the NSF hopes to make up for some of the expected shortfalls. Roger Blandford, director

of the Kavli Institute for Particle Astrophysics and Cosmology in Stanford, California, chairs the panel that will recommend where the cuts should fall. The Washington meeting was the last of seven held in an attempt to placate astronomers around the country since the panel was announced in August (see *Nature* **436**, 616; 2005).

Blandford says the panel will consider the scientific value of the various facilities, as well as their training capabilities, decommissioning costs and overlap with other observatories. Scientific value, in particular, is notoriously difficult to measure quantitatively — an attempt by one astronomer at the conference demonstrates as much (see 'Which sites get cited?', opposite). "There are no easy targets," says Blandford. "We are trying very hard to make this a responsible process."

It was clear that many of the 200 or so astronomers in the room were not convinced. "If we go through with the cuts then major facilities will disappear," warns Jeffrey Linsky, an astronomer at the Joint Institute for Laboratory

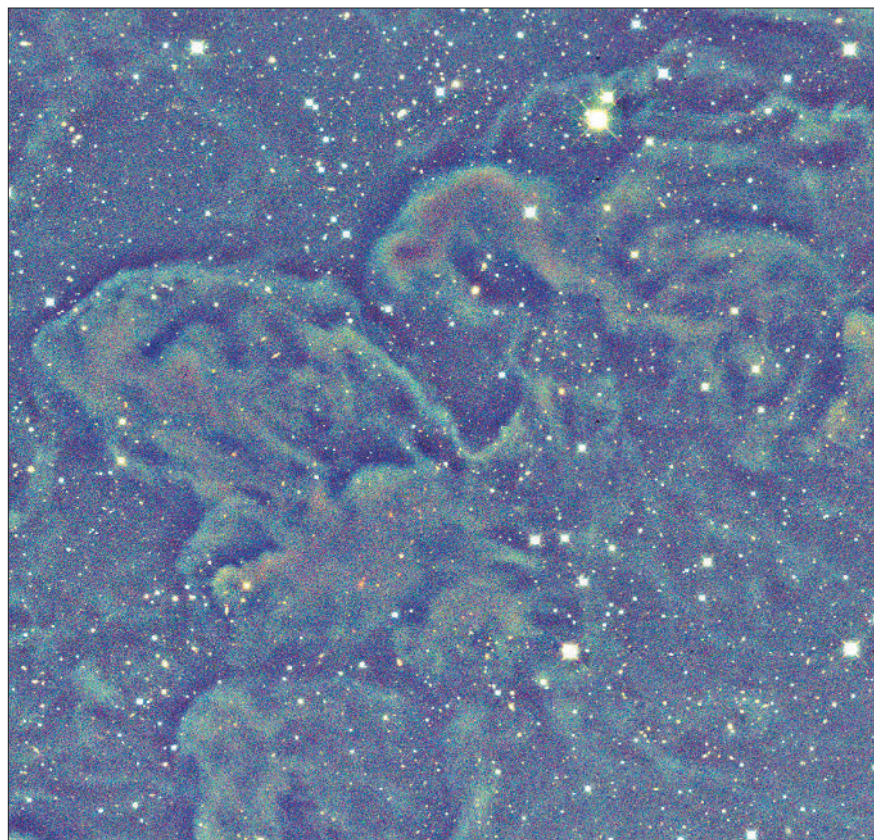
"I say we complain, and complain bitterly."

SNAPSHOT

Cloudshine is a stellar snap for Harvard duo

These ghostly swirls are the first images of gas clouds reflecting starlight. Lying more than 1,000 light years away, they are part of the Perseus Molecular Cloud, a region where new stars form. Usually, such clouds appear in astronomical images merely as dark regions where light from more distant stars is blocked. Thermal emission from the gas and dust can often be detected by radio telescopes, but directly reflected starlight — 'cloudshine' — is normally too dim to see.

Jonathan Foster and Alyssa Goodman of the Harvard-Smithsonian Center for Astrophysics in Cambridge, Massachusetts, captured these false-colour images of the Perseus complex using the new OMEGA 2000 camera at Calar Alto Observatory in southern Spain (J. B. Foster and A. A. Goodman *Astrophys. J. Lett.* **636**, L105–L108; 2006). They say that scattered light alone can explain the images, and that the near-infrared cloudshine could even be used to probe the three-dimensional structure of molecular clouds. ■



J. FOSTER & A. GOODMAN/HARVARD-SMITHSONIAN CENTER FOR ASTROPHYSICS

Astrophysics in Boulder, Colorado. "I say we complain, and complain bitterly."

Van Citters told the crowd that this is not an option. "We have to go through this," he said. Blandford agrees: "It is economically necessary, but it is also politically necessary." Without proving that astronomers are serious about reining in spending, they are unlikely to win fresh funding from the White House.

Blandford says the committee has received many letters and e-mails since the review was

announced, most defending the merit of individual facilities. He and his panel members will now deliberate, aiming to produce recommendations by the end of March.

At least astronomers have no cause to take the measures personally. In an era of flat budgets, such panels are likely to become commonplace, predicts Van Citters. "I suspect within a year or two we're going to see more of this throughout the foundation."

Geoff Brumfiel

Which sites get cited?

Which astronomical observatories are the most productive? The question is bound to provoke arguments, as it depends on the definition of productivity. But Virginia Trimble, an astronomer and expert in scientometrics at the University of California, Irvine, has taken a stab at a quantitative answer.

Trimble combed through more than 3,500 astronomy papers published in 2001, separating them by wavelength and noting which telescopes, satellites and balloon instruments supplied the data. If more than one observatory contributed data, she divided the credit evenly. Her intern Paul Zaich then went to the Science Citation Index and tallied the number of citations for each paper over the subsequent three years, yielding a citations-per-paper index for each observatory.

Their results, reported at last week's meeting of the American Astronomical Society in Washington DC,

are sometimes surprising. Big telescopes produce more papers than small ones, as might be expected, and their papers are more often cited.

But even though the Hubble Space Telescope led the 250 or so optical and infrared



telescopes with 16% of the papers and 19% of the citations, Trimble says it was not particularly highly cited per paper: "It's close to the average." The optical facility that produces the most highly cited papers, although not vast numbers of them, is the Sloan Digital Sky Survey.

The Very Large Array in

New Mexico (pictured) dominates in the radio spectrum even more than Hubble does in the optical, accounting for 22% of radioastronomy papers and 27% of citations. But Australia's Parkes telescope had a higher citations-per-paper index (20.52 versus 16.55). The most highly cited space-based observatories other than Hubble were the two newest: XMM-Newton and the Chandra X-Ray Observatory.

Many factors can influence the results, from an observatory's location to different practices among astronomers in citing each other's work. So the arguments won't end, and scientists whose telescopes don't score highly will still grumble. One reviewer even took Trimble to task for having too informal a writing style. But, she says, the survey is done partly for fun. "If I can't tell any jokes at all, this isn't worth doing."

Tony Reichhardt

Selected astronomical facilities and impact factors

	Papers	Citations	Citations per paper
Maxima balloon experiment	5.6	456	81.43
Sloan Digital Sky Survey	31.3	1,252	40.00
Chandra	175.8	6092	34.60
Anglo-Australian Telescope	45	1,253	27.84
Keck	104.5	2,180	20.86
Very Large Array	181.4	3,003	16.55
William Herschel Telescope	44	686	15.59
Hubble Space Telescope	346.3	4,747	13.71
Very Large Telescope	69.8	864	12.38

Source: Virginia Trimble, Paul Zaich, Tammy Bosler

ON THE RECORD

“He’s attractive. And my son loves the fact he was an astronaut.”

A deep-thinking Canadian voter describes why she might vote for Liberal parliamentary candidate Marc Garneau.

“The Americans are going to be last on the list, I can tell you.”

A group of researchers at Britain’s Mill Hill lab discuss how they will share their samples after US news coverage played down the fact that they were the ones to isolate and grow the Turkish bird-flu virus.

Sources: *Montreal Gazette, The Independent*

SCORECARD



Extreme tourism
For just US\$334 a person, Ukrainian travel agents offer an all-inclusive holiday to Chernobyl to mark the 20th anniversary of what they call the “worst environmental disaster in history”.



Scientific vanity
Go one better than Googling yourself: new features on Elsevier’s Scopus service make it easier to get daily updates on who is citing your work — and that of your competitors.



Plant patents
Local farmers are calling on the University of Hawaii to give up its patents on three kinds of taro, the root that forms the basis of much Hawaiian native cuisine and culture.

NUMBER CRUNCH

NASA’s Stardust mission returned to Earth last weekend bringing with it samples snared from a comet (see page 255). It was quite a trip:

4.6 billion kilometres were travelled by Stardust on its journey to Comet Wild 2 and back.

7 years is the time it took the probe to do the round trip.

Less than 0.1 gram is the estimated amount of cometary particles collected by Stardust.

Journals submit to scrutiny of their peer-review process

Results are starting to come in from what may be the largest-ever study of the practice of peer review.

Three prestigious medical journals — *The Lancet*, *Annals of Internal Medicine* and the *BMJ* — threw open their doors and allowed the study team unprecedented access. Researchers were able to see all relevant paperwork, including confidential referee reports, and were permitted to video editorial meetings. The result is a data set that documents the passage of more than 1,000 papers through the peer-review process, from submission to publication or, far more often, rejection (see ‘The data set’).

Having cameras in meetings was a bizarre experience, says Richard Horton, editor of *The Lancet* — he says it felt like the researchers were making a reality television show about medical journals. But Horton adds the process was well worth it, as “this is a huge study, which makes the results very reliable”.

Those results, the first of which were submitted for publication earlier this month, give a qualified thumbs up to current editorial practices. They may also go some way to dispelling widely held doubts about peer review.

One question investigated by the study authors, who are based at the University of California, San Francisco, is whether editors favour positive results over null findings. Many researchers do not attempt to publish negative results, assuming that editors are not interested. This is a particular problem in medicine, because the efficacy of a drug will be exaggerated if trials reporting negative results are not published.

But when the California researchers

The data set

1,107 papers were examined
100+ hours of meetings were taped for each journal
16 editors were interviewed

compared the 68 manuscripts that were published with those rejected, they found no evidence of bias towards statistically significant results. “Hopefully, this will encourage authors to submit,” says Kirby Lee, an expert in evidence-based health care and an author on the study. “Publication bias is a serious problem; it can really skew results in meta-analyses.”

Other findings also tend towards the quantitative (see ‘Dissecting peer review’). But peer-review experts say the qualitative parts of the study, which had to wait for transcripts of the videos, are likely to prove more interesting. “Peer review can be a complex decision-making process involving lots of people,” notes Sara Schroter, who studies journal practice at the BMJ Publishing Group in London. “If you want to understand why something happens it is best to conduct qualitative research.”

Lee says that the qualitative study should shed light on issues such as the criteria that editors use when deciding to review or reject articles, as well as how interactions at editorial meetings help shape decisions about publication. “We want to identify sources of systematic bias in the editorial review process that may result in a publication record that is not representative of the true distribution of study findings submitted to each journal,” he says.

Horton adds that the study could also help improve public understanding of peer review. The timing is good, he says, because questions about editorial standards are being asked in the wake of the scandal surrounding the South Korean stem-cell scientist Woo Suk Hwang, who published two widely acclaimed papers later found to have been faked. Horton says that some criticisms, such as failure to spot fabricated data, stem from a lack of understanding about what peer review can and cannot do.

“Peer review is a black box to the public and politicians,” he says. “Unless we open up that box we are going to get misperceptions.” ■

Jim Giles

Dissecting peer review

What the researchers have discovered so far.

- There is no right answer — referees often disagree with each other.
- Authors have to come clean — they frequently fail to disclose funding sources and potential conflicts of interest in submitted manuscripts until asked to do so by journal editors.
- It separates the wheat from the chaff — the methodological quality of accepted articles is higher than that of rejected ones.



CLIMATE CHANGE IN FOCUS

Find all our news on global warming in one easy place.
www.nature.com/news/infocus/climatechange.html

Promises to clean up industry fail to convince

SYDNEY

Some of the world's biggest polluters and energy consumers met last week under a scheme trumpeted by organizers as a "complement" to the Kyoto Protocol. Members of the Asia-Pacific Partnership (AP6) on Clean Development and Climate promised to provide practical solutions to climate change, by driving industrial partnerships and encouraging new, cleaner technologies. But with voluntary participation, no emissions targets, no deadlines and little new money, environmental groups are somewhat sceptical. Can industry really be counted on to clean up its act?

The inaugural meeting of the AP6, held in Sydney on 11 and 12 January, was attended by senior government and business representatives from Australia, China, India, Japan, South Korea and the United States — countries that together are responsible for around half of the world's greenhouse-gas emissions.

But environmental groups are not impressed by AP6's unwillingness to set targets or adopt carbon trading. "It's incredibly disappointing that AP6 isn't prepared to put in place financial mechanisms to reward those who invest in cleaner energy and penalize those who don't," says Erwin Jackson of the Australian Conservation Foundation, based in Melbourne.

Many industry representatives see it differently. "I don't think the first step should be to create an unequal playing field by putting carbon-trading mechanisms in one place and not in another," counters Oscar Groeneveld, chief executive of the Rio Tinto Aluminium group.

So the meeting focused instead on garnering a range of softer commitments from industry — to share knowledge, develop technology and improve operating practices. The main achievement was the establishment of eight government-industry task forces to



Protesters express scepticism while representatives of the world's top polluters meet in Sydney to consider cleaner ways.

focus on power generation; coal mining; building and appliances; the production of cleaner fossil energy, renewable energy, steel, aluminium and cement. The groups will meet separately to formulate priorities, action plans and progress indicators, before reporting back to the next AP6 gathering, probably in January next year.

Supporters insist this will bring significant gains. "There are some

low-hanging fruit," says Groeneveld. "If we lift the performance of the 'bottom half of class', we can improve the whole industry and substantially reduce emissions."

Both the Australian and US governments committed new funds: AU\$100 million (US\$75.5 million) over five years, and a one-off US\$52 million respectively. But industry is largely expected to foot the bill. Details of exactly what will change are lacking, however, as industry groups say they will essentially continue their ongoing investment into research and development.

Advocates of AP6, including US Secretary of Energy Samuel Bodman, point to the aluminium industry as evidence that the voluntary approach can work. The industry set a series of "voluntary objectives" in 2003, such as reducing perfluorocarbon (PFC) emissions, which Groeneveld claims

have been "collectively reduced by 75% since 1990".

That doesn't impress the critics. "There is not much new money on the table and we already have a large number of technology-transfer mechanisms in place," says Iain McGill, an engineer who is researching energy markets at the University of New South Wales in Sydney.

"You have to question how significant the initiative is," says McGill. He and others are sceptical that things will change fast enough without government regulation. "Even if you develop the most whiz-bang technologies, you still have to get them in place," he says. "It might happen — but you wouldn't want to

bet the climate on it."

"The scale of the climate changes being projected for 2050 are so substantial that to say, 'there will be a technological fix' is inadequate," adds Andy Pitman, a climate scientist at Macquarie University in Sydney. "As scientists, we haven't managed to get across the urgency of the problem."

Even the petroleum giant BP seems unconvinced by the voluntary approach. "Low-emission technologies are available now," points out a spokesman for BP Australia. But to ensure their uptake "market pull is essential," he says, adding that to "reduce the costs of low-carbon technologies to parity with conventional power sources".

Perhaps this incentive, if unspoken, was felt at the AP6 meeting. "If industry doesn't act responsibly, governments will have to intervene and regulate," says John White, chairman of the Perth-based company Global Renewables. "No one wants to talk about it, but that was definitely a take-home message." ■
Carina Dennis

STRINGER/REUTERS

Stress makes medics ever gloomier

NEW YORK

Did you wake up this morning feeling blue? If you're a medical researcher, the answer may well be yes. A survey of faculty members in US medical schools has revealed that as many as one in five report signs of depression.

Grumbles about spiralling stress are rife among researchers, but there have been few studies to gauge its effects in any field. So psychiatrist Barbara Schindler at Drexel University College of Medicine in Philadelphia and her team sent questionnaires to more than 3,500 faculty members in four US medical schools. The questions were part of established scales for measuring mental and physical health, as well as life and job satisfaction.

The team received responses from more than half of the academics, and published its results this month (B. A. Schindler *et al. Acad. Med.* **81**, 27–34; 2006). The most striking finding is that some 20% of researchers, both male and female, show symptoms consistent with

clinical depression. This is roughly double the rate of such symptoms seen in the general population, and is a jump from the 14% found in a similar study in 1984.

Younger researchers show more depressive symptoms than older ones, and basic researchers feel less strain than those who also see patients. Although the survey is small, the authors say it exposes the toll that work is taking on academics' mental health. "When 20% of doctors are depressed you know something is wrong," Schindler says.

Researchers report various reasons for growing strain at work. Faculty members at US medical schools say they are under increasing pressure to see fee-paying patients. This leaves them less time to pursue the research and publications that win promotion. "I think these things mean there is less happiness and relaxation in medical schools," says George Mandel of George Washington University

Medical Center in Washington DC.

The repercussions are harder to predict. The Association of American Medical Colleges (AAMC) in Washington DC reports no fall in the number of faculty members it can recruit,

"When a fifth of doctors are depressed, something is wrong."

or in the numbers dropping out. But increasing stress could be dissuading faculty members from carrying out research, says Kenneth Getz, whose work

at the Tufts Center for the Study of Drug Development in Boston suggests that the number of principal investigators leading clinical trials has declined in the past few years.

Diane Magrane, who is responsible for faculty development at the AAMC, says that many medical schools are already aware of poor staff morale and are introducing mentoring and support programmes to combat it. "When you measure it, it allows you to do something," she says. ■

Helen Pearson

German science receives promised boost in funds

The new German government is on course to keep its pre-election pledge to bolster public investment in research.

Chancellor Angela Merkel announced on 10 January that an additional €6 billion (US\$7 billion) will be injected into basic and applied research over the next four years. As long as private industry also pitches in, the extra money would boost Germany's research spending to 3% of its gross domestic product by the end of 2009.

How exactly the money will be spent, and which fields of science may benefit, will be agreed by research minister Annette Schavan and other government departments, probably by the end of February. The cash is also likely to allow Schavan to run a planned €1.9-billion competition for elite universities without cutting funds elsewhere.

DNA tests confirm guilt of executed prisoner

Roger Keith Coleman was guilty after all. New DNA tests prove that in 1981 Coleman raped and murdered his sister-in-law in Grundy, Virginia, a crime for which he was executed in 1992.

Many opponents of the death penalty had hoped the tests would prove that Coleman was innocent. This would have made him the first person in the United States to be cleared by DNA evidence after their execution (see *Nature* 439, 126–127; 2006).

Outgoing Virginia governor Mark Warner had requested the tests be done on old DNA samples taken from the scene of the crime. Researchers from Canada's Centre of Forensic Sciences in Toronto reported the results on 11 January.

"We are very disappointed," says Kate Germond of Centurion Ministries, a non-profit organization based in New Jersey. An exoneration would have boosted the anti-death-penalty cause. "It would have been tremendous," she says.



Groups opposed to the death penalty were dealt a blow by DNA tests on an executed prisoner.

Researchers reach out for a little bit of stardust

Some 180 scientists across the world are gearing up to receive tiny amounts of the cometary particles collected by NASA's Stardust mission.

The spacecraft (shown here during re-entry) parachuted safely onto the Utah desert on 15 January, carrying thousands of dust particles collected from the tail of Comet Wild 2. Stardust gathered the particles on an aerogel paddle about the size of a tennis racquet. The dust was destined for NASA's Johnson Space Center in Houston before being shared out among researchers.

Scientists hope that the samples will give them clues about the chemical make-up of the primordial rubble that spawned the planets.



Stardust's close encounter with the comet, in January 2004, has already revealed that Wild 2 has a surprisingly rigid core and a surface that is pock-marked by craters.

Google Scholar search engine goes multilingual

Chinese- and Portuguese-speaking scientists can now search the scientific literature in their own languages using the free search engine Google Scholar. Additional languages, including French, Spanish and German, will be added later.

Abstract and indexing databases such as PubMed, Scopus and Thomson Scientific's Web of Science already include English abstracts of foreign-language journals, where such information is available. But Google Scholar allows full-text searching of some articles in their original language, ranking papers by the number and quality of citations.

Anurag Acharya, the engineer who built Google Scholar, says that Chinese and Portuguese are the first languages to be added because there were local journal databases to collaborate with in these countries, including Wanfang Data and VIP Information in China, and SciELO and Bireme in Brazil.

US drug agency gives green light to early clinical trials

The US Food and Drug Administration (FDA) issued new guidelines on 12 January to pave the way for more drugs to make the leap from lab bench to patient.

The regulations allow scientists to make small batches of trial drugs in their labs instead of ordering them from factories. Researchers can then test smaller doses in fewer patients than those used in early clinical trials. They will use imaging studies to gauge whether the drug reaches its intended target before embarking on a full-blown clinical trial. The agency has been working to distinguish promising drugs earlier in the development process, as nine out of ten experimental drugs fail

when they are tested in humans.

Andrew von Eschenbach, the acting FDA commissioner, reassured watchdog groups that the new guidelines would not place patients at risk. "Rapid does not mean reckless," he said.

Private donation set to keep collider running

The Relativistic Heavy Ion Collider, the largest US nuclear-physics experiment, has received a \$13-million private donation that will allow it to keep operating this spring.

The collider, which smashes atomic nuclei together to recreate conditions in the early Universe, was facing a budget shortfall in 2006, according to Praveen Chaudhari, director of the Brookhaven National Laboratory in New York, where the collider is based. The Department of Energy had given the lab enough money to run the accelerator for only five weeks — just long enough to warm it up and cool it down again. Under those circumstances, Chaudhari says, "we were not going to make a run" this year.

The donation comes from James Simons, mathematician and head of investment management company Renaissance Technologies, based in Setauket, New York. The directors of Brookhaven Science Associates, which runs the lab, and of which Simons is a member, are also involved in the donation.

"It seemed a terrible shame that so valuable a piece of scientific equipment and so valuable a team of scientists be left for a year to lie fallow," Simons says.

Correction

The 5 January News Feature 'Awash with fossils' (see *Nature* 439, 14–16; 2006) misidentified palaeoanthropologist Berhane Asfaw as director of Ethiopia's National Museum. He is a former director of the museum; the current director is Mamitu Yilma.

A SEA CHANGE

A collapse in ocean currents triggered by global warming could be catastrophic, but only now is the Atlantic circulation being properly monitored. **Quirin Schiermeier** investigates.

Henry Ellis, captain of the British slave-trader *Earl of Halifax*, had a scientific bent. While sailing the subtropical Atlantic in 1751, he measured water temperatures at different depths, using a thermometer, a long rope and a bucket fitted with flaps that sealed water inside the vessel when it was raised. Ellis was surprised to find the coldest water in a mid-ocean layer around 1,200 metres below the surface. The Sun, he concluded, did not warm the ocean in proportion to depth.

The discovery proved useful for Ellis's crew: "By its means we supplied our cold bath, and cooled our wines or water at pleasure," he wrote in his notes. But the global significance of the Atlantic's cold depths escaped Ellis and pretty much everyone else for the next two centuries.

He had stumbled upon the generator of a world-girdling system of currents — an enormous flow of water known as the 'global conveyor belt'¹, which transports warm surface water towards the poles and cold deep water back to the tropics. Driven by differences in temperature and salinity, this 'thermohaline' circulation has in recent years become infamous as the possible cause of major climatic upheaval. But only in the past year have much-needed automated systems been installed to monitor this circulation almost constantly. "There's a crying need for these data," says Gavin Schmidt, a leading climate modeller at the NASA Goddard Institute for Space Studies in New York. "For the first time we'll be able to observe the ocean 'weather' in all its complexity."

The cold water Ellis had found in the Atlantic's depths comes from two regions at the ocean's north end, in the Greenland and Labrador seas. Here, saltier water coming northwards cools and sinks, before reversing south. This great submarine U-turn is peculiar to the waters of the North Atlantic, whose extreme cold temperatures and saltiness give it a higher density than is found in the Indian and Pacific oceans.

Evidence from the ice ages suggests that shifts in the thermohaline circulation have dramatic effects on the temperature in western Europe and beyond; past shutdowns of



the conveyor drastically cooled the climate all around the North Atlantic in a matter of years by stalling the currents that bring warm water northwards². And computer models suggest that, in a seeming paradox, intense regional cooling could be triggered by global warming³. By the beginning of this century, the apparent fragility of the thermohaline circulation had made it by far the best-known exemplar of the surprising, non-linear and potentially catastrophic shifts in climate that makes the prospect of a greenhouse world so scary.

Current affairs

But the flows themselves remain surprisingly unmeasured. Until this year, almost all attempts to monitor what is happening in the Atlantic's depths have relied on some form of Captain Ellis's method — roaming along the surface and dredging up water from various depths as one goes. This year, scientists will have access to continuous measurements collected by 22 moored 'profilers' — sensors that travel up and down wires from buoys to moorings on the sea floor taking measurements as they go. The profilers were set up last year by a UK programme called Rapid Climate Change (RAPID), a £20-million (US\$35 million), six-year programme of the Natural Environment Research Council which has installed these profilers as part of a wider scheme to quantify the likelihood and magnitude of rapid climate change in the future.

Climatologists worldwide are anxious to get hold of these data. The most recent shipboard study, published in *Nature* last year, suggested that the circulation might be yet



Harry Bryden's research suggests that Atlantic currents have slowed in recent years.

ALAMY

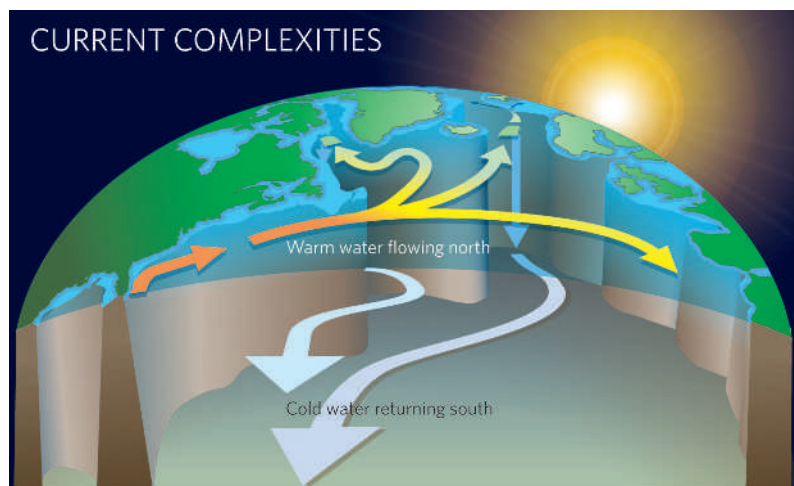
RAPID

more fragile than had been thought⁴. But at the same time, other research suggests that its potential to do harm may be much subtler than images of a Europe thrown into a mini ice age suppose.

The idea that changes in ocean circulation might be a key determinant of climate change dates back to the early twentieth century and to the great American geologist Thomas Chamberlin. In the 1950s, the oceanographer Henry Stommel pioneered scientific understanding of the three-dimensional structure of the Earth's oceans, and of the currents that flow one way on the surface and another way at depth. But the theorizing that brought the North Atlantic branch of the great conveyor to its present fame dates back only to 1984, when Wallace Broecker, a geochemist at Lamont Doherty Geophysical Observatory at Columbia University in New York, attended a talk in Bern by Hans Oeschger, a Swiss climatologist. While Oeschger outlined his latest findings about climate instabilities and large oscillations of atmospheric carbon dioxide during the most recent ice age, it occurred to Broecker that a switching on and off of the thermohaline circulation in the North Atlantic could be the missing link. Temporary failure of the Atlantic conveyor could have wreaked havoc on climate, he thought.

Although the carbon dioxide fluctuations Oeschger wanted to explain later proved to be artefacts, the idea that the conveyor could stop and start with planet-juddering effects took off. In 1985, Broecker and his colleagues published a landmark paper⁵ drawing on early computer models of the ocean's flow. They proposed that the Atlantic circulation had two distinct stable modes — one with the conveyor on and one with it off — and that it was relatively easy for it to move from one mode to the other. The distinction between the two modes, they suggested, might explain the difference in climate between ice ages and warmer interglacials. Soon thereafter, computer models began to show that an increase of carbon dioxide in the atmosphere might, by increasing the temperature and thus the supply of fresh water to the North Atlantic, cause just such a shutdown.

The idea of a threshold that, if passed, could cause calamity, or as Broecker termed it, “a nasty surprise in the greenhouse”, has played an increasingly important role in predicting the consequences of a greenhouse effect. In the late 1990s, William Calvin brought the idea to a wider audience with his article entitled ‘The great climate flip flop’, which graced the cover of *The Atlantic* — as a neurophysiologist, Calvin had been interested in whether rapid climate change had been a decisive factor in human evolution. A few years later, a 2003 report for the Pentagon, ‘Imagining the unthinkable’, described how rapid climate change caused by such a shutdown could pose threats to whole societies and the peaceful coexistence of nations. Shortly thereafter, a film called *The Day after Tomorrow* pictured the citizenry of the United States chased over the Mexican border by an instant ice age; again, the North Atlantic was to blame.



“For the first time we’ll be able to observe the ocean ‘weather’ in all its complexity.”
— Gavin Schmidt



Among the first to see the effect of the global conveyor on climate was Wallace Broecker.

Given the thermohaline circulation's pivotal role in discussions of climate change, there was much excitement when, last November, *Nature* published evidence suggesting that the system could have slowed down dramatically⁴. The evidence had been gathered by Harry Bryden, an oceanographer at the University of Southampton, UK, and his team, while on a research cruise that also put the finishing touches to the deployment of RAPID.

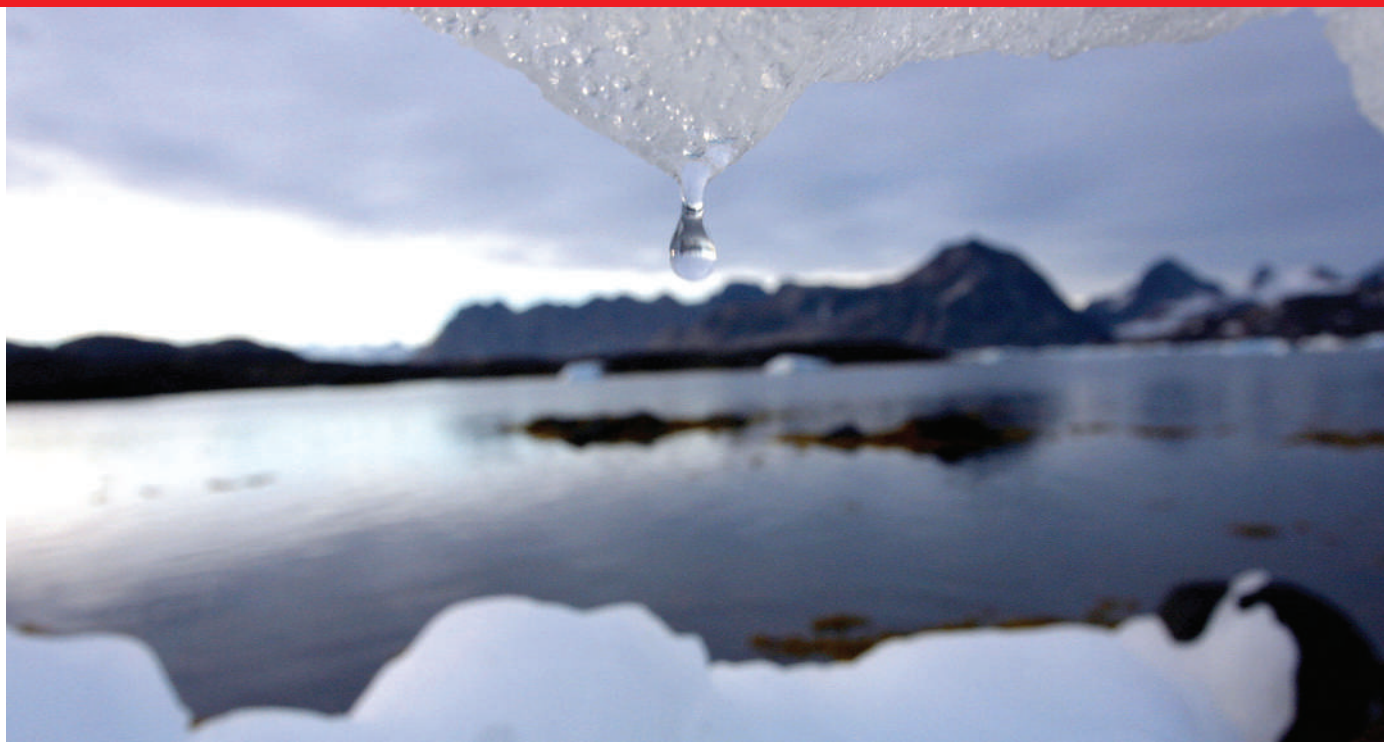
Comparing their 2004 measurements with data from 1957, 1981, 1992 and 1998, Bryden and his colleagues found that some of the warm surface water that used to flow northwards now seemed to remain trapped in the subtropical Atlantic, looping east and then returning south rather than heading north. Altogether, the ‘overturning’ circulation at 25° N — the latitude where Ellis had first probed the ocean 250 years before — seemed to have decreased by about 30%.

In too deep

The result came as a surprise to those in the field. Few scientists had thought that such dramatic slowing of the thermohaline circulation could happen so soon. Models suggest⁶ that the increase in fresh water needed for a conveyor shutdown would not be expected without a global warming of 4–5 °C; warming in the twentieth century is currently put at 0.6 °C (ref. 3). The most complex computer models of the climate and oceans, the sort used to make climate predictions for the Intergovernmental Panel on Climate Change (IPCC), suggest that the flow might be expected to slow by an average of 25% by the end of the twenty-first century, but not to shut down completely³.

Running complex models long enough to simulate some sorts of change, however, uses an unfeasible amount of computing power. So for some purposes ‘intermediate’ models can capture things better. Stefan Rahmstorf, an oceanographer at the Potsdam Institute for Climate Impact Research (PIK) in Germany recently compared the circulation's response to an influx of fresh water in 11 simpler models; all showed a threshold, called the bifurcation point, beyond which the thermohaline circulation cannot be sustained⁷. The size of the threshold suggests that the possibility of shutdown is real, but not immediate. Rahmstorf says, “It is very unlikely that it will become really critical for the thermohaline circulation within the next 100 years.”

This is not to say that freshwater flows are not increasing; they are. The annual runoff of the six mightiest rivers draining into the Arctic Ocean, including Russia's Ob, Lena and Yenisey, is now 128 cubic kilometres greater than it was when routine measurements began 70 years ago⁸, an



increase of about 7%. In addition, rising temperatures are making sea ice melt more rapidly. Perhaps most important, the huge Greenland ice sheet is showing worrying signs of disintegration; it is currently thought to be shrinking by 50 cubic kilometres per year⁹.

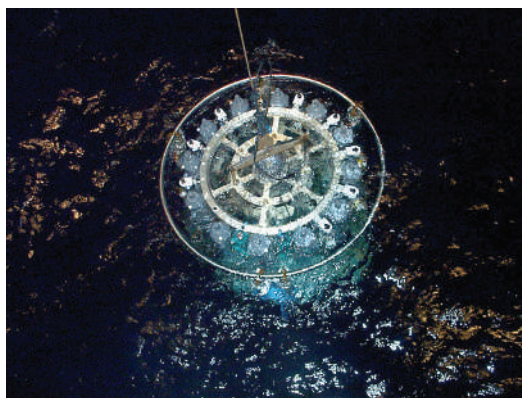
Ruth Curry, an oceanographer at Woods Hole Oceanographic Institution in Massachusetts, has investigated how much of this extra fresh water lingers in the parts of the Greenland and Labrador seas that are critical for the functioning of the thermohaline circulation. Her recent analysis¹⁰ of 1950 to 2005 salinity data suggests that 4,000 cubic kilometres — eight times the annual outflow of the Mississippi river — of fresh water have accumulated in the upper ocean layers since the 1960s. “The extra freshwater input is beginning to affect density,” she says. But the amount of fresh water needed to shut down the thermohaline circulation in Rahmstorf’s comparisons is an order of magnitude greater than the flux reported by Curry, and she agrees that the circulation will not be unduly affected this century. Peter Wadhams, an oceanographer at the University of Cambridge, UK, last year reported a substantial weakening of convection ‘chimneys’ down which surface water flows in the Greenland sea, but it is unknown how much of the observed effect is due to natural variability.

This is all hard to reconcile with Bryden’s findings, which suggest that a strong slowdown is already under way. “Something strange is going on here,” says Michael Schlesinger, a climate modeller at the University of Illinois at Urbana-Champaign who views the possibility of a thermohaline circulation shutdown as more likely and more worrying than many of his peers. “If Bryden’s findings are real it means that the circulation is much more sensitive to fresh water than any model has ever predicted.”

It is not just that the results are unexpected —

Warming is increasing the flow of fresh water into the sea, which could trigger the collapse of currents.

Profiles of salinity and temperature can now be obtained from sensors moored in the Atlantic ocean.



they also seem hard to reconcile with other data. If the circulation were slowing down as Bryden suggests, one might expect that Europe would already be getting colder. The North Atlantic transports around a petawatt of heat — equivalent to the thermal output of about 500,000 large power stations — towards Europe. Interrupting that flow should have a cooling effect on the climate, but no such change has been seen.

A fragile balance

It may be that the system has a previously unexpected level of natural variation. Or it could be that Bryden recorded noise, rather than a signal — did a set of readings, through coincidence, the presence of ocean eddies and other natural disturbances, make it seem that the circulation was slowing when it wasn’t? A statistical artefact cannot be excluded. “The results are based, after all, on just five snapshots of an extremely noisy and under-sampled system,” says Carl Wunsch, a physical oceanographer at the Massachusetts Institute of Technology, who harbours long-standing doubts about the significance of the thermohaline circulation and its possible shutdown. “The story is appealing, but it is a very extreme interpretation of the data. It’s like measuring temperatures in Hamburg on five random days and then concluding that the climate is getting warmer or colder.”

In response to his critics, Bryden points to data on the density of the ocean at various depths gathered at the same time as the flow readings, which provide independent support for the idea that the circulation is slowing. But although other scientists are less harsh than Wunsch, many remain cautious. “Bryden’s results are extraordinary,” says Schmidt, “but this is exactly why they also require extraordinary evidence.”

If Bryden’s results are correct, there is another explanation of the lack of cooling in Europe: that a slowdown of the thermohaline circulation will not have the dire effects that have been suggested. It may be that, in today’s climate, the role of the thermohaline circulation in warming Europe has been overestimated. A paper published in 2002 suggested that the westerlies, the dominant winds in mid-latitudes that blow from west to east play a much larger role

FOX-WARNER



The shutdown of the Atlantic currents plunged New York into an ice age in *The Day After Tomorrow*.

than was long thought¹¹. But much of the heat transported in the atmosphere ultimately comes from the ocean. “It is true that the atmosphere does the heavy lifting,” says Jeff Severinghaus, an oceanographer at the Scripps Institution of Oceanography in La Jolla, California, who was once a student of Broecker’s. “But the ocean exerts the control, just like the driver of a car.”

Evidence for the huge effects of past thermohaline shutdowns is near indisputable. The best case is that of a 1,300-year cold period that occurred around 12,000 years ago, known as the Younger Dryas. The carbon isotope ratios in fossilized plankton from the period suggest that the thermohaline circulation was much slower than it is today (slow circulation allows light carbon isotopes to build up near the ocean’s surface).

This slowdown coincided with a vast surge of fresh water into the North Atlantic. The melting of the ice-caps as the ice age ended created a vast reservoir of fresh water known as Lake Agassiz. It was far larger than any of today’s Great Lakes, over parts of Minnesota, Dakota and Manitoba — Lake Agassiz. To the east, the lake was bordered by a tongue of the Laurentide ice sheet. When the tongue collapsed, a huge amount of water flooded down the St Lawrence River and into the Atlantic.

According to ice cores drilled in Greenland, similarly large temperature oscillations — the Dansgaard-Oeschger events that first piqued Broecker’s interest in the 1980s — took place throughout the 90,000 years of the most recent ice age. It is likely that they were also caused by the thermohaline circulation stalling.

But in this respect, as in others, the past may not be a straightforward guide to the present. The consequences of a shutdown could depend on the climate at the time the current stalls. Broecker now believes that the cooling in the Younger Dryas and the Dansgaard-Oeschger events came about because the shutdown of the thermohaline circulation was exacerbated by a positive feedback, in the form of enhanced winter sea-ice formation. An influx of fresh water at high latitudes encourages the formation of sea ice, because fresh water freezes more easily. Because ice reflects sunlight, and stops heat from the ocean below reaching the atmosphere, spreading sea ice would strongly amplify cooling due to thermohaline slowdown, especially in winter. Studies of moraines in Greenland and Scandinavia show that during the Younger Dryas the cooling in summers was relatively moderate, whereas in wintertime temperatures must have been more than 30 °C lower than now.

It is hard to evaluate the amplifying role of sea ice very precisely. Most coupled ocean-atmosphere models include a sea-ice component, but the representation is crude and leads to an unrealistic simulation of sea-ice distributions. If this feedback is as important as Broecker thinks, then the effects of a thermohaline circulation shutdown in a warmed world may be very different from those seen during the ice ages and their immediate aftermath. Today, satellite images show sea-ice cover at a

historic low. In a world that had undergone the degree of warming needed to trigger a thermohaline shutdown in most models there would be almost none.

Rahmstorf speaks for many climate researchers when he rejects the idea that a thermohaline shutdown in today’s climate would lead to a rerun of the Younger Dryas, in which large parts of Europe were frozen. “You can’t just assume a linear relationship and say that everything will happen on a 5° higher level,” says Rahmstorf. Broecker still believes that global warming may have surprises in store, possibly including a collapse of the thermohaline circulation, but he agrees that “the notion that it may trigger a mini ice age is a myth”.

Earth watch

The fact that a future shutdown might not have the predicted effects on climate might go some way to explaining how Bryden could observe the circulation slowing — or at least fluctuating — without major climatic consequences, at least so far. Although Severinghaus agrees that this may be part of the story, he and many of his peers would rather believe that there was a randomly wrong signal in the data. “It just doesn’t quite fit,” says Schmidt. “If the circulation has been 30% down for a decade, it should at least have produced a 1–2° drop in sea surface temperature even if it didn’t cool Europe. But no such thing has been observed.”

Bryden says that the new RAPID system for monitoring flow in the Atlantic should allow them to know within a decade whether they found a long-term slowdown or a natural fluctuation. Other new approaches may also help. The ARGO system, part of the international Global Climate Observing System, is a fleet of robotic floats that monitors temperature, salinity and current in the upper 2,000 metres of the Indian, Atlantic and Pacific oceans. The free-drifting floats sink to pre-established depths and then surface to transmit their data to satellites. ARGO data are invaluable for monitoring changes in remote ocean regions, according to Lynn Talley, a physical oceanographer of the Scripps Institution of Oceanography in San Diego, California. For example, they have already revealed a spectacular warming of the southern ocean surrounding Antarctica, she says.

And *in situ* monitors are not the only way of keeping an eye on the deep ocean. A weakening of the thermohaline circulation would change the entire topography of the sea surface, says Rahmstorf. Such large-scale changes could be picked up by satellites. A recent simulation¹² suggests that the sea level of the North Atlantic could rise locally by up to a metre as a result of adjustments to the density flows below the surface; in some regions the rate of change could be up



Carl Wunsch believes the effect of Atlantic circulation on climate has been overblown.

to 2.5 centimetres per year. Scientists have begun using satellite altimetry to see if such changes are already observable; again, they expect robust results within a decade.

Modellers also have much to do. Most model studies, such as those used by the IPCC, look at how a freshwater-induced shutdown of the thermohaline circulation might change temperatures if everything else remained the same. A harder question is what a shutdown might mean in a world that is, on average, getting warmer. Bryden's findings have caused a stir throughout the climate research community; lead authors of the chapters on ocean physics and circulation in the IPCC's fourth assessment, due in 2007, are reworking their submissions.

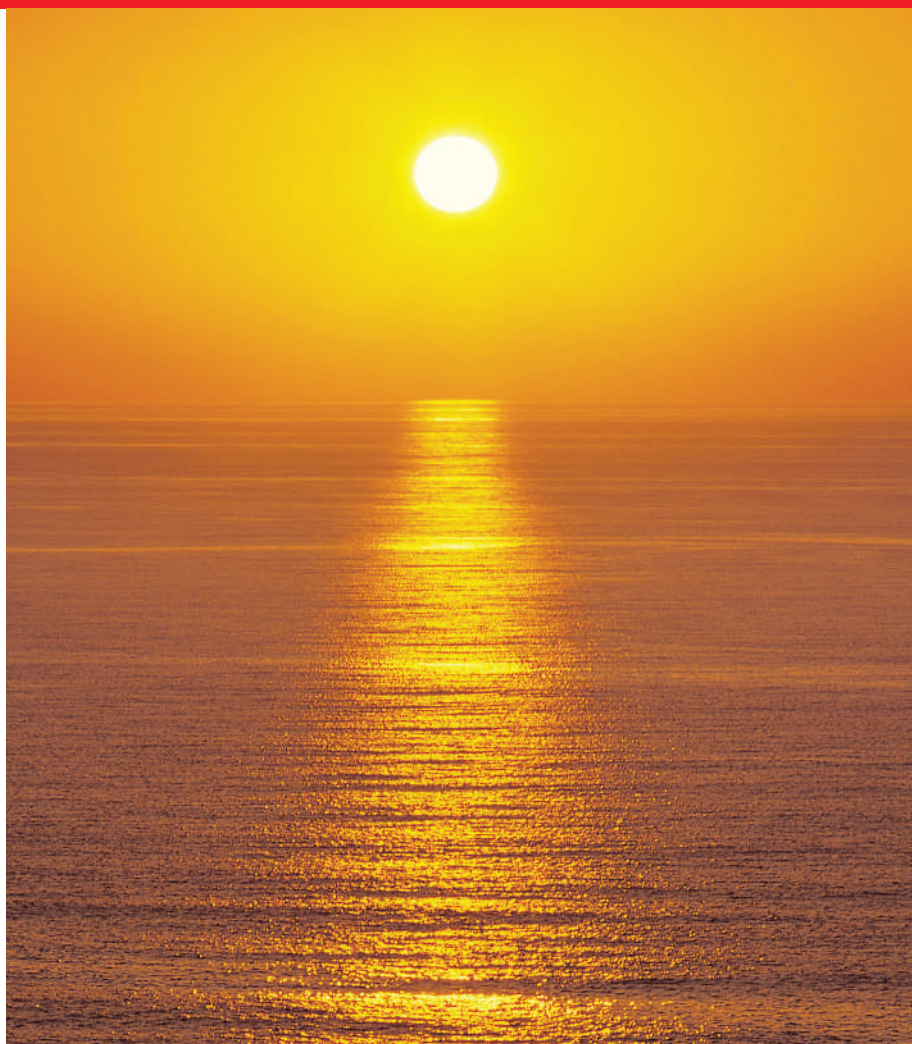
Future unknown

Wolfgang Cramer, an ecologist at PIK, predicts complex changes in the climate, with some effects exacerbating each other and some that cancel each other out. For example, Cramer says, meteorological perturbations caused by a thermohaline shutdown could lead to a dramatic increase in the frequency of major floods and storms in large parts of Europe even if overall temperatures do not drop. "It's not the mean, it's the extremes that are most worrying," he says.

One aspect of the problem is that the thermohaline circulation is not just a climatic affair. Its effect on ocean circulations means it influences the rates at which nutrient-rich bottom water rises to the surface all around the world. A recent simulation suggests a shutdown might lead North Atlantic plankton stocks to collapse to less than half their current biomass¹³. Globally, a decline of more than 20% might be expected thanks to reduced upwelling of nutrient-rich deep water and gradual depletion of upper-ocean nutrient concentrations.

"Plankton builds the base of the marine food web. So a decline in global plankton biomass and productivity can be expected to have consequences for fish, squid and whales as well," says Andreas Schmittner, a climate researcher at Oregon State University in Corvallis. "A weaker Atlantic overturning circulation could result in a reduced fish supply to people living along the shore lines of the Pacific and Indian Oceans."

Other possible effects of a shutdown predicted by models include warming in the tropics, or, rather surprisingly, over Alaska and Antarctica. Rainfall patterns might change, too. A southern shift of the thermal equator — which has accompanied thermohaline circulation shutdowns during ice ages — could lead to monsoon failures, and droughts in Asia and the Sahel region, says Severinghaus, and these effects seem to be independent of sea ice. Such shifts could have severe consequences for poor farmers in many parts of the world, consequences that may be considerably more disruptive than colder winters in affluent northern Europe, says Severinghaus. And, as Schlesinger points out, a weakening or stopping of the thermohaline circulation would reduce the carbon dioxide uptake of the ocean, which would mean a positive feedback on global warming. The



oceans currently absorb about a third of the carbon dioxide released from fossil fuels, although the proportion is set to decrease as emissions climb.

Some 250 years after Captain Ellis first probed the Atlantic, its depths still hold secrets and threats. Even in a new age of constant monitoring and improved modelling, it will be some time before the likelihood, and the probable effects, of a thermohaline circulation slowdown can be predicted with accuracy. The intricacies of a system that depends on delicate balances between fresh and salt water over vast ocean basins, on the details of atmospheric circulation, wind-driven currents and the topography of deep sea floors will not yield answers quickly. "If you would like to learn how a planet operates you would probably not choose the Earth," remarks Schlesinger. We greenhouse dwellers, alas, do not have a choice. ■

Quirin Schiermeier is *Nature's* German correspondent.

"The notion that a collapse of the thermohaline circulation may trigger a mini ice age is a myth."
— Wallace Broecker

1. Broecker, W. S. *Oceanography* **4**, 79–89 (1991).
2. Broecker, W. S. *Nat. Hist. Mag.* **97**, 74–82 (1987).
3. Intergovernmental Panel on Climate Change *Climate Change 2001: The Scientific Basis* (Cambridge Univ. Press, Cambridge, 2001).
4. Bryden, H. L., Longworth, H. R. & Cunningham, S. A. *Nature* **438**, 655–657 (2005).
5. Broecker, W. S., Peteet, D. M. & Rind, D. *Nature* **315**, 21–26 (1985).
6. Stocker, T. F. & Schmittner, A. *Nature* **388**, 862–864 (1997).
7. Rahmstorf, S. *et al. Geophys. Res. Lett.* **32**, doi:10.1029/2005GL023655 (2005).
8. Peterson, B. J. *et al. Science* **298**, 2171–2173 (2002).
9. Schiermeier, Q. *Nature* **428**, 114–115 (2004).
10. Curry, R. & Mauritzen, C. *Science* **308**, 1772–1774 (2005).
11. Seager, R. *et al. Q. J. R. Meteorol. Soc.* **128**, 2563–2586 (2002).
12. Levermann, A. *et al. Clim. Dynam.* **24**, 347–354 (2005).
13. Schmittner, A. *Nature* **434**, 628–633 (2005).

Hooked on fossils

For decades, much of the early history of fish evolution was locked away in rocks in China.

Rex Dalton tracks down the scientist who brought many of the remains to the surface.

The dry desert setting of Mesa, Arizona, may not seem the most appropriate place to talk about the watery world of fish. But palaeontologists gathered in a lecture hall there last autumn to celebrate a life spent studying ancient oceans and the fish that swam in them.

The symposium, held by the Society of Vertebrate Paleontology, celebrated one of China's most prominent palaeontologists — Meemann Chang, former head of the Institute of Vertebrate Paleontology and Paleoanthropology (IVPP) in Beijing. Her work has helped clarify the links between the fish that swam in Earth's oceans 400 million years ago and the air-breathing, land-walking creatures that evolved from them. Now 69, she has played a key role for years in bringing little-known Chinese fish fossils to the attention of the scientific world.

Chang's career has been far from straightforward. Along with her successes, she has also faced significant difficulties, thanks to the shifting political landscape of her homeland. Despite such problems, her enthusiasm for her subject remains undimmed. "I am still digging and collecting fossil fishes," she smiles during an interview at the landlocked Arizona hotel.

Chang's father, a gifted pathologist from Nanjing, wanted her to become a physician, but love of her country led her to choose geology instead. In 1958, during the Great Leap Forward, she was among those who heeded the call of vice-president Liu Shaoqi to study the Earth so that China might exploit its natural resources, such as oil. For Chang, that introduction to China's rocks set her on the path to study fish fossils, a quest that has taken her to all the continents of the world.

In 1965, Chang was chosen to do graduate research at the Swedish Museum of Natural History in Stockholm, one of the leading research centres in palaeontology. But her time there was to prove short-lived. When the Cultural Revolution swept China in 1966, Chang, ever the patriot, halted her studies and returned home. In Beijing, Chang was



Chipping away: Meemann Chang has devoted her life to finding and understanding fossilized fish.

confronted by the new phenomenon of the Red Guard who, on the orders of Mao Zedong, 'purified' China by isolating and punishing the academic classes.

Bad dream

For more than a decade, Chang lived what is now called the 'time of nightmares' — public humiliations to challenge the intellectual spirit, and hard labour in the countryside to break the body. "I wasn't allowed to do research," she recalls, "only to read Mao." It would be many years before Chang was able to return to Sweden to complete her doctorate.

"She is a wonderful person who has been through a lot," affirms John Maisey, curator of fossil fish at the American Museum of Natural

History in New York and one of the organizers of the Mesa meeting. "But she still smiles and is charming."

Chang's career has taken her through many countries, and allowed her to pick up numerous languages. She earned her undergraduate degree in 1960 at the Lomonosov Moscow State University, where she became fluent in Russian. She learned modest Swedish while in Stockholm, is fluent in English and reads German and French. But she is also adept at deciphering another language: that of fossilized remains¹. She can readily navigate a path from the 'age of fish' 400 million years ago in the Devonian period, through to the end of the dinosaur age and the Cretaceous period 65 million years ago.

In her current studies, Chang is working to understand the species distribution pattern of fish across the Pacific Ocean — a distribution that reached its maximum during the Eocene epoch, between 34 million and 56 million years ago. Most of these fish became extinct in the western Pacific, she notes, but a few, such as the coelacanth, still survive in the eastern Pacific. "Tracing the origins and distribution of these fish is a very exciting endeavour," she says.

Chang's contribution to Chinese palaeontology was recognized in 1983 when she became the first woman to head the IVPP. This was significant not only because of her gender



Missing link: remains of the fish *Youngolepis* have helped plug gaps in the evolutionary tree.

M. CHANG

B. CHOO

but because it marked the IVPP's move away from political appointments to those based on merit. Chang served two terms as director, ending her tenure in 1990, and helped shepherd the institute from the days when whole families were living on an upper floor of the research building, to a new facility that included modern laboratories.

Unlike some of her more rigid compatriots, Chang was very flexible and open when it came to guiding her students' careers, says palaeontologist Desui Miao, who helped to organize the Mesa symposium. He cites the case of Zhonghe Zhou as an example. One of Chang's promising students, Zhou had begun a doctoral programme in the early 1990s to examine fossil fish. But then quarries in the northeastern province of Liaoning started to yield an intriguing assortment of fossilized birds dating back to the early Cretaceous.

Zhou saw this as an opportunity to switch from fish to avian fossils. Chang agreed, allowing him to change the direction of his research. "This was a major break with Chinese tradition," says Miao, of the University of Kansas in Lawrence. "But it showed how she treated every student," he adds — working first and foremost to develop them professionally.

The change more than paid off. What began as a seemingly minor academic move helped pave the way for China to become a leading force in palaeontology. Liaoning's avian-like fossils of feathered dinosaurs with rapacious teeth redefined how birds evolved². Soon, the world's top palaeontologists were clamouring to come to China, which in turn generated collaborations and opportunities abroad for young Chinese researchers.

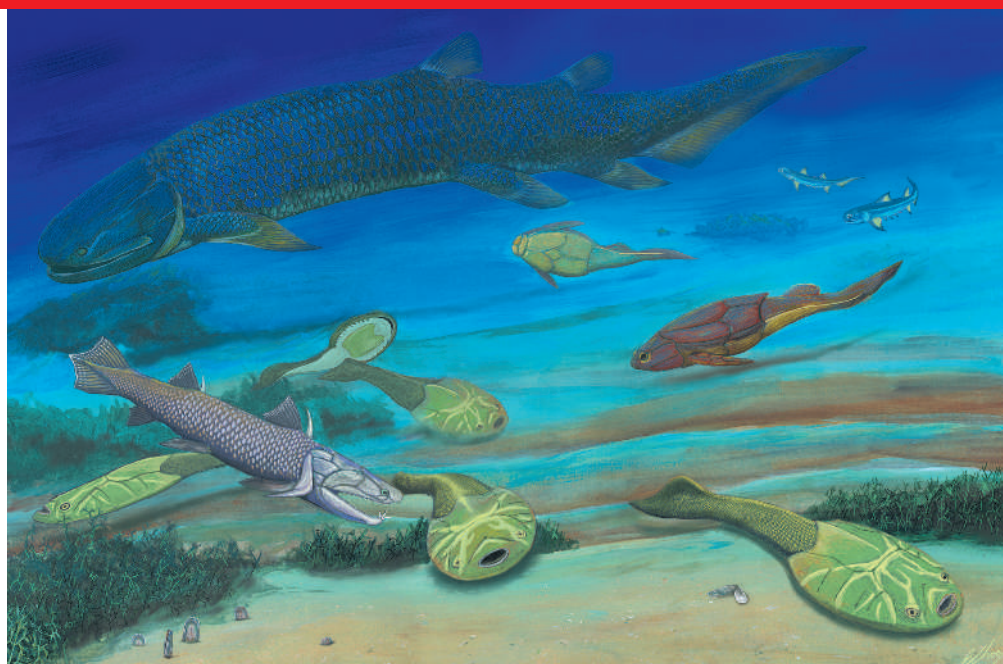
Miao himself was among those who benefited from these new links — in the late 1980s, he found himself studying at the University of Chicago. Once more, Chang was to show her willingness to put her students' interests first. In 1988, Miao knew he wanted to continue his postdoctoral studies in Chicago, but at that time China's leaders, worried about a country-wide 'brain drain', were pushing for foreign-trained scientists to return home.

Taking a gamble

Miao decided to write to Chang, asking her permission to stay in the United States. A Chinese colleague thought this was a rash move, calling him a "bloody fool", Miao remembers. But soon after, Miao received a letter from Chang granting her permission. "I was stunned," he recalls. "For the first time in a long time, I wept."

Chang's experiences in Stockholm, of course, meant that she knew only too well the difficulties of studying abroad. But she also understands the rewards. Despite the interruption by the Cultural Revolution, her research in Sweden did much to rework the evolutionary tree for fish — and sparked some very lively debate.

Hans-Peter Schultze, a palaeontologist who was doing a postdoc at the Stockholm



Reconstruction of fossilized fish found in the Yunnan province, including *Youngolepis* (upper left).

museum in the early 1960s, remembers the rumours of fabulous specimens from the early Devonian that Chang had brought from quarries in Yunnan province. At the time, palaeontologists regularly argued about the evolutionary tree of fish before species evolved to move ashore. Such trees, or cladograms, are important in understanding historical biodiversity and specialized characteristics of current species.

Swedish palaeontological icons Erik Stensiö and Erik Jarvik — both now deceased — held strongly to a view about the split between two

using cladograms to challenge her conclusions about how closely *Youngolepis* and *Diabolepis* are related to the dipnoans.

Lars Werdelin, a graduate student in Stockholm when Chang returned to complete her graduate degree, says her understated manner made her data even more convincing. "She doesn't stretch the evidence," says Werdelin, now senior vertebrate curator at the Stockholm museum. "She is not prone to hyperbole. When she says something, you believe it."

Although colleagues often tell of Chang's personal warmth, they acknowledge that she also has a steely, determined side. In the late 1950s, she was a student leader charting a future field trip in a dangerous area of Kazakhstan. Outsiders were loathed then, and hotels used to deny foreigners a room. "She demanded a room — arguing, patting her side and saying: 'I have money. I have money,'" says Ke-Qin Gao, a palaeontologist at Peking University in Beijing, who heard the story later. "She was fearless." And she got the room.

Today, Chang never tells such stories. Asked about her successes, she brushes aside the questions, seeking credit for her students and colleagues.

Fortunately, her students and colleagues have found a way to honour her record. Xiaobo Yu, a palaeontologist at Kean University in Union, New Jersey, is preparing a book based on the Mesa symposium. Yu couldn't go to college during the Cultural Revolution. But afterwards, Chang took him on as her first graduate student. It set him on a course to receive his doctorate from Yale University. ■

Rex Dalton is Nature's West Coast correspondent.

What began as a seemingly minor academic move helped pave the way for China to become a leading force in palaeontology.

lineages of Devonian fish: lungfish (dipnoans) and lobe-finned fish (porolepiformes). Before Chang's work, there was no known species that shared characteristics from both these types of fish, which were the predecessors of creatures that later walked on land and breathed air. But Chang had a fossilized fish that did: *Youngolepis*, a specimen dating from around 415 million years ago in the Devonian³.

"It was very disturbing for them when Chang brought the new form," says Schultze, who is now at the University of Kansas. "Jarvik called *Youngolepis* the 'devil's fish'." In jest, Chang later used that epithet to name another Chinese specimen, *Diabolepis*, which furthered her theories of the link between lungfish and lobe-finned fish. Her specimens "became pivotal in strengthening the connection" between these species, says Schultze, and helped to lay the groundwork for Chang to propose an evolutionary history for the fishes. Debate over this history continues today, with some authors

1. Chang, M. *Nature* **403**, 152–153 (2000).
2. Hou, L.-H., Zhou, Z., Gu, Y. & Zhang, H. *Chinese Sci. Bull.* **10**, 61–63 (1995).
3. Chang, M. in *Origins of the Higher Groups of Tetrapods: Controversy and Consensus* (eds Schultze, H.-P. & Trueb, L.) 3–28 (Cornell Univ. Press, Ithaca, New York, 1991).

BUSINESS

Hwang scandal hits Korean biotech hard

Until its ugly end, 2005 had been a pretty good year for South Korea's fledgling biotech industry. The government was pouring money into biomedical research, and the Kosdaq stock market brought in changes that made it easier for biotechnology companies to go public. Many firms saw their share prices surge.

But they hit the buffers in December, when the magnitude of the scandal consuming Woo Suk Hwang, a stem-cell biologist at Seoul National University, became apparent. Early in the month, stock value crashed even for companies whose business is not directly related to stem cells, such as MacroGen (see graph), which provides tools and services for genome research. Most firms lost between a fifth and a half of their value in December alone.

The industry was so badly shaken partly because Hwang's research — much of it now discredited — had played a starring role in its ascent, insiders say. "Hwang provided the psychological effect" behind the earlier growth, says Han Oh Park, chief executive of Daejeon-based Bioneer, which makes reagents and drug-discovery tools.

At the same time, the South Korean government's generous financial support for Hwang's work, which has totalled perhaps US\$30 million, boosted investors' confidence that biomedical research in the country would yield significant results. Thanks largely to Hwang, more biotech companies expressed a specific interest in stem-cell research — and when they said as much, their stock prices rose.

Jung Seob Shin of KDB Capital, a venture-capital arm of the Korea Development Bank, argues that last month's collapse does not reflect a change in the fundamental outlook of the sector. He says that this has been



LEE J.-M./AP

The fall of Woo Suk Hwang has damaged even companies that have nothing to do with stem cells.

improving over the past few years, and points out that in the first half of this month, many stocks have already regained ground.

Se Jong Park, president of the Korea Bio Venture Association in Seoul, goes so far as to argue that the scandal could help the industry's long-term growth, by reminding investors to look more closely at individual stocks. "What will happen during 2006 is discrimination" between good and bad companies, he predicts. Of about 600 biotechnology companies in South Korea, Park estimates that just 50 will survive in the long term.

And there are still signs of confidence in Korean biotechnology. Three companies — Seoul-based gene-therapy company Viomed, Bioneer and CrystalGenomics, which makes platforms for protein crystallization — went public in late December or early January; they raised a total of US\$72 million between them. "I was very worried about the impact of the scandal, but we overcame it," says Bioneer's Park.

Not everyone can expect to be so fortunate. "Fundraising will become more difficult for certain companies," says Jason Lim, who manages investment in biotech for InterVest, a venture-capital company in Seoul.

The most immediate concern for the Korean biotech industry is that the government, which increased its annual spending on biomedical research from some 380 billion won (US\$380 million) in 2001 to more than 700 billion won last year, might now divert resources into rival spheres of research, such as computer science or nanotechnology. For now, the ministry of science and technology says it has no plans to slow its investment in biomedical research.

Ichiko Fuyuno

IN BRIEF

MONOPOLY MONEY Boeing and Lockheed Martin look set to form a joint venture that could dominate the supply of rocket launchers to the US military. Officials close to the deal say that the Pentagon has endorsed the venture and will shortly take it to the US Federal Trade Commission, which regulates competition, for approval. Rivals are aghast that the Department of Defense is prepared to let its two largest rocket suppliers work together in this way — but analysts expect to see more such deals between defence contractors as spending is squeezed.

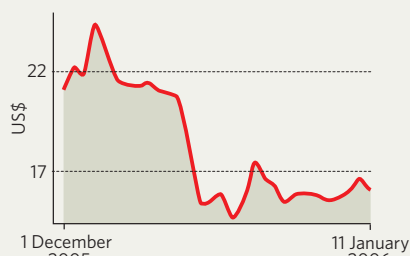
RECORD REVENUES

Biotechnology company Genentech clocked revenues of \$6.6 billion in 2005, capping a year of extraordinary success fuelled by cancer-drug sales. The South San Francisco firm reported that its profits in the last quarter of the year grew by 64% compared with 2004. It also scored a public-relations coup by topping *Fortune* magazine's list of the 100 best US companies to work for.

WILL IT SHELL OUT? Royal Dutch Shell is being sued by 26 European pension funds, mainly based in the Netherlands, over its admission that it artificially boosted its claimed oil and gas reserves between 1997 and 2003. The pension funds are claiming hundreds of millions of dollars in compensation from the oil company; they have broken away from litigation already being undertaken in the United States.

DEAL SEALED Shareholders in Berna Biotech, a Swiss vaccine-maker, voted last week to approve a takeover by Dutch biotechnology company Crucell (see *Nature* 438, 737; 2005). The merger resulting from Crucell's all-stock offer, valued at about US\$460 million, will create the world's largest independent vaccine-maker; the company will retain the name Crucell and stay based in Leiden. The merger was approved hours after Novartis dropped an effort to acquire Berna — late last month Novartis said it was considering buying the smaller company and combining its vaccine operations with those of its expected acquisition, California-based Chiron.

MACROGEN STOCK



Risks of resurrecting 1918 flu virus outweigh benefits

SIR — Your Special Report “The 1918 flu virus is resurrected” (*Nature* **437**, 794–795; 2005), on the debate surrounding the recreation of the 1918 (Spanish) human influenza virus, focuses solely on the question of publication, giving little attention to the possibility of regulating potentially harmful experiments before they are started.

A key element of co-operative security is trust and confidence. Only a high level of transparency will enable nations to judge others' intentions correctly, act to discourage unfounded suspicion and build confidence in compliance with arms-control treaties.

Withholding critical information on dual-use research would undermine arms-control efforts. Once a research project has been conducted, it should be published in full. The preferred option for research where potential risks are considered to outweigh potential benefits is to stop the research before it even starts. Both the UK Royal Society and the US National Academy of Sciences have proposed a systematic review of scientific research proposals before work starts, but this approach is entirely neglected in the debate about the Spanish flu virus. The consequences of this project could, and should, have been assessed ten years ago or at any time since.

Another shortcoming of the current debate is the lack of a systematic approach to risk–benefit analysis. Any intervention in the scientific process should be based on a scientifically sound assessment of both risks and benefits. Criteria to assess the potential misuse of dual-use research include its military usefulness and the level of technical expertise needed to apply this knowledge for malign purposes, taking into account a variety of potential actors. The availability of risk-management tools, such as effective arms-control measures, is also an important factor in risk assessment.

Although most commentators agree that the misuse potential of the Spanish-flu work is comparatively high, it seems that the objective assessment of benefits is the more contentious part of the equation.

Your Special Report claims that the Spanish-flu work increases understanding of virulence and pathogenicity factors, and might contribute to identifying the next pandemic strain or developing appropriate drugs. These general statements hold true for most biomedical research projects. But generalities contribute little to a systematic risk–benefit analysis, which should include an assessment of the importance of the research in a health and humanitarian context, whether alternative research avenues are available and the added value of any particular experimental approach.

Influenza pandemics are an important

public-health problem, but it is questionable whether a reconstructed viral strain from 1918 is necessary to address this problem. Hundreds of other influenza strains from the past five decades, varying highly in terms of contagiousness and pathogenicity, provide an abundant research resource for comparative studies. The added value of one additional strain — and thus the concrete benefit of the reconstructed Spanish flu strain — is very limited.

Dr Jan van Aken

Hamburg Centre of Biological Arms Control,
Hamburg University, Falkenried 94,
20251 Hamburg, Germany

Value of high-protein diet is clearer than drawbacks

SIR — There may be disagreement about whether the words “scientifically proven” should be used to sell books (“A recipe for trouble” *Nature* **438**, 1052; 2005). However, we have published widely on the subject of high-protein diets and our findings are broadly similar to those of other research groups who have shown better health outcomes on this kind of diet (see, for example, D. K. Layman *et al.* *J. Nutr.* **133**, 411–417; 2003, and A. Due *et al.* *Int. J. Obes. Relat. Metab. Disord.* **28**, 1283–1290; 2004).

High-protein diets have been criticised for their potential to cause renal and bone disease (J. Eisenstein *et al.* *Nutr. Rev.* **60**, 189–200; 2002) and the red-meat component has been linked to colorectal cancer (A. Chao *et al.* *J. Am. Med. Assoc.* **293**, 172–182; 2005), but the evidence is contradictory.

As *The CSIRO Total Wellbeing Diet* was for a general readership and covered a range of topics, we did not report an extensive review of the literature on high-protein diets. Our own work has focused not just on reduction of body fat but on reduction of lipids, glucose, insulin and blood pressure and minimization of lean-tissue loss. In our opinion, the high-protein, moderate-carbohydrate approach is superior in all these respects. We do not agree with the view reported in your News story (“Diet book attacked for its high-protein advice” *Nature* **438**, 1060; 2005) that the CSIRO diet could lead to more breast cancer and prostate cancer.

You discuss the support we receive from the Australian meat and dairy industries. These industries funded two out of the five weight-loss studies we have performed with this protocol. These protocols were investigator-devised and controlled; the funding bodies had no input into the reports and papers. Certainly they capitalized on the positive results, as is their right.

Peter Clifton

CSIRO Human Nutrition, PO Box 10041 BC,
Adelaide, South Australia 5000, Australia

Saintly helpers at hand in Renaissance hospital art

SIR — In his *Science in Culture* article (“A vision of birth” *Nature* **438**, 1084; 2005), Martin Kemp comments on medical allusions in the nativity scene painted by Hugo van der Goes in the fifteenth century. Religious iconography was very important at that time, and the saints present in the painting deserve further comment.

The female figure on the far right with an elaborate white dress represents Saint Elisabeth of Thuringia, a thirteenth-century Hungarian princess who was one of the earliest known founders of a hospital in Europe. She stands next to Mary Magdalene, and is carrying a vase of myrrh, a symbol of healing, which I believe symbolizes the early nursing sciences.

On the far end of the left panel is Saint Anthony the Abbot, founder of the first monastic orders, which ran all hospitals at the time. Next to him, carrying a lance, stands Saint Thomas Didymus (“Doubting Thomas”), who was renowned for his insistence on evidential proof.

These saints made excellent icons for the Renaissance healthcare professionals: doctors, friars and nurses. They are portrayed by van der Goes as equal, in size and importance, to Mary and Joseph and at least twice as large as the donors pictured below.

Piero Dolara

Department of Pharmacology, Viale Pieraccini 6,
University of Florence, Italy

Humour of gene names lost in translation to patients

SIR — The choice of a gene name can have unforeseen consequences in addition to infringement of trademark (“Pokémon blocks gene name” *Nature* **438**, 897; 2005). The quirky sense of humour that researchers display in choosing a gene name often loses much in translation when people facing serious illness or disability are told that they or their child have a mutation in a gene such as *Sonic hedgehog*, *Slug* or *Pokemon*.

As with the acronym CATCH22 (from ‘cardiac anomaly, T-cell deficit, clefting and hypocalcaemia’) for chromosome 22q11.2 microdeletions, which was abandoned because of its no-win connotations (J. Burn *J. Med. Genet.* **36**, 737–738; 1999), researchers need to be mindful when naming genes and syndromes.

Ken Maclean

Developmental Biology Unit,
Victor Chang Cardiac Research Institute,
384 Victoria Road, Darlinghurst,
New South Wales 2010, Australia

COMMENTARY

Experiments in social responsibility

Pursuing drugs for neglected diseases is not a traditional part of the pharmaceutical company portfolio. But **Paul Herrling** of Novartis finds that it brings welcome changes both within and outside the industry.

Pharmaceutical companies are commercial enterprises, and until recently were almost exclusively focused on generating maximal returns for their shareholders. We are now seeing more of these companies invest their hard-won returns in providing greater access to medicines for patients in poorer parts of the world. These projects have a distinctly charitable aspect to them and will not generate profits (see table overleaf for examples). The most recent initiative by the Swiss company Novartis, for which I work, is the creation of an institute in Singapore (the Novartis Institute for Tropical Diseases; NITD) dedicated to the discovery of drugs for tuberculosis and dengue. I believe that these activities are generating positive changes within our industry, and also in our relationships with our partners in these initiatives — non-governmental organizations (NGOs) and other stakeholders — who are usually among our harshest critics.

In the past, most NGOs active in health care in the developing world considered the pharmaceutical business as part of the problem rather than a potential partner in any solutions. From my personal perspective, relations have greatly improved since Novartis established its 'access-to-medicines' projects. For example, the Global Alliance for Tuberculosis Drug Development (GATB) has sought our expertise in industrial drug discovery and has nominated me to their board. Yves Champey of the Drugs for Neglected Diseases initiative (DNDi), an offshoot of the medical aid agency Doctors Without Borders (or MSF), has also invited me to join their scientific advisory committee.

On the defensive

MSF used to be uniformly critical of the pharmaceutical industry, but now has a more heterogeneous attitude. Some individuals remain negative, as demonstrated by MSF's public criticism of Novartis when we ran into supply problems last year with the raw material for our malaria medicine Coartem, which we provide at cost to patients in Africa. Still, others within MSF have become partners and regular visitors to our Singapore institute, and



The new Novartis Institute for Tropical Diseases in Singapore holds bright hopes for drug discovery.

we are grateful for their willingness to teach us what they know about dengue and tuberculosis patients in Asia and Africa (see below).

Although many public-health NGOs remain highly critical of intellectual-property protection, both the DNDi and GATB engage, to different degrees, in 'defensive' patenting to protect medicines they helped to discover and develop. Defensive patenting means pursuing legal protections that allow them to do what they like with new drug compounds, including the right to make them available at cost without having to pay licences or royalties to any third parties, such as universities or biotech and pharmaceutical companies. Defensive patenting would even allow them to generate a return on investment should those medicines find customers in wealthy markets — a return they could use to finance further research and development into neglected diseases.

As well as the changing attitudes of NGOs, changes have also occurred within our industry. Projects on neglected diseases at Novartis have required adjustments in the way that drug-discovery scientists and other personnel operate. For example, to develop medicines it is imperative to understand precisely the patients' problems, their environment and the doctors treating them, as well as having the necessary access to patients or their tissues.

These are issues we understand well for the diseases of richer societies but less so for tropical diseases. In particular, we had little prior experience with patients affected by tuberculosis and dengue, which affect hundreds of millions of people worldwide each year.

Lessons to be learnt

Our first decision, and not a completely trivial one, was where to locate our new institute. Our existing research centres in Europe and the United States fulfilled all criteria for excellence except one: proximity to patients and their doctors. This need drove a wider search that led us eventually to southeast Asia and Singapore. When we began our research activities at the NITD we were able to visit hospitals in neighbouring Cambodia, Vietnam, Indonesia and India to familiarize ourselves with conditions there. In addition, we visited tuberculosis and dengue patients in sub-Saharan Africa, where conditions differ from those in southeast Asia.

Our first lesson was the need for good, cheap and easy-to-use diagnostic methods, which are lacking for both diseases. Second, we learnt that treatments must not only be cheap to make, but simple to use in the field or in the relatively primitive conditions of district hospitals in these parts of the world. Any treatment requiring sophisticated equipment would not be viable, ruling out many 'high-tech' medical tools used by richer societies.

Third, we learnt that many prospective patients in these regions have a different understanding of medicine from us, and are

"In the past, most NGOs active in health care in the developing world considered the pharmaceutical business as part of the problem rather than a potential partner in any solutions."

not familiar with our 'scientific' approach to it. Most patients, particularly in rural regions, typically go to traditional healers first and have a mistrust of Western medicine. They will often only go to hospitals or health centres when traditional methods have failed. Sometimes this means their condition has deteriorated to a point that might have been avoided had they sought medical help earlier.

Patients in tropical regions are much more embedded in their local community than are Western patients, and consequently are best reached not as individuals but through the community. So, when introducing new therapies, it is important to gain the support of community leaders first and to communicate our medical concepts in ways that make sense within that culture. Experience shows that once this is achieved, patients in Africa, for instance, are actually more compliant in following a treatment regimen than are Western patients, who may second-guess their doctors and stop taking recommended medicines.

Wider considerations

Importantly, we learnt the significance of reinforcing infrastructure at locations where we wanted to conduct research or clinical studies, long before the new medicines are available. We also learnt to evaluate any potential corruption issues affecting an investment site,

which are not only incompatible with our corporate ethics but also risk distorting scientific results, or placing researchers in danger. We were fortunate to find many hospitals and research centres that have dedicated and skilled local personnel who are capable of delivering impressive results in extremely difficult conditions. One example is the Ifakara Center in Tanzania, which has successfully conducted malaria clinical studies using good clinical practice standards. We are in the process of identifying similar sites in Africa and Asia and building relationships for research and clinical development. This is time-consuming and requires intensive investment in forming personal relationships.

For drug-discovery personnel, who are used to working for a large pharmaceutical company focused on patients in richer societies, this was a crucial learning experience for achieving their new goals on neglected tropical diseases. The final task was something we had to unlearn: to actively suppress our usual criteria for short-term commercial viability and accept that these projects will not have any financial returns.

Why have the attitudes of some pharmaceutical companies and their shareholders changed from exclusively seeking profits to a limited, but significant, support of access-to-medicine activities? Such investments would

have been much rarer only 20 years ago, but today our company directors and senior management actively encourage us to pursue this work. The wider public is also supportive, and our shareholders have so far not been critical.

Why the change?

I cannot fully explain this attitude shift, but if I allow myself to speculate, one factor is probably genuine compassion. With improvements in communication technologies, including 24-hour television coverage of natural disasters, the plight of patients in poor regions has never been so evident to the citizens of rich countries. Sometimes this coverage gets translated into aid, as evidenced by the tremendous success of the donation campaigns for the 2004 tsunami in Asia.

A further, more pragmatic factor might be the realization by wealthy societies that it is unwise to ignore the diseases of poorer countries, as these diseases can circle the globe with increasing speed through travel and tourism. We cannot easily forget that HIV originated in Africa and SARS reached Canada from southeast Asia in a matter of weeks. Today, the perceived threat of a flu pandemic arising from avian flu in Asia brings back memories of the devastating 1918 'Spanish' flu. It seems likely, therefore, that richer societies will increasingly ask their pharmaceutical companies to allocate some resources to diseases of poor societies so as to develop global defences against these potential threats.

A third factor, arising from within the industry, is the knowledge that pharmaceutical companies are currently seen by many outsiders in a negative light and that some companies would like to change this perception. This, in turn, improves morale within the company.

Of course, there are limits to what we can achieve. The prime mission of a commercial enterprise remains the generation of returns for investors, so only limited resources will be allocated to not-for-profit activities. And it is true that only commercially successful companies can afford such initiatives. But it is also the case that for some neglected diseases a fairly modest investment can have a huge impact.

Such not-for-profit activities would be impossible without personal commitment from the highest level of management. At Novartis, we are fortunate that our chairman, Daniel Vasella, is a medical doctor interested in access-to-medicine issues. He has encouraged us to think of contributions we could make in this area and consistently supported the solutions we proposed. These activities are new experiments for Novartis, and although we do not know how successful they will be, or how long they will be supported by our shareholders, we hope they will encourage others to do the same — to the benefit of all patients. ■

Paul Herrling is head of corporate research at Novartis International, CH-4002 Basel, Switzerland.

Work on neglected diseases by pharmaceutical companies and their partners (see www.ippph.org)

Company	Project	When	Where	Partners
GlaxoSmithKline	R&D on oral malaria treatments	2000 and ongoing	Sub-Saharan Africa	World Health Organization (WHO)
GlaxoSmithKline	Albendazole donations to combat elephantitis	1998 and ongoing	12 countries in Africa	WHO, Bill & Melinda Gates Foundation, Merck, national health ministries
GlaxoSmithKline	Allocation of laboratory resources for neglected diseases	2003 and ongoing	Madrid, Spain	MMV for malaria, GATB for tuberculosis
Merck	Donation of ivermectin to combat river blindness	1987 and ongoing	Countries in Africa, the Middle East and Latin America	See www.mectizan.org/partners.asp
Merck	HIV/AIDS partnership	2004	Botswana	Bill & Melinda Gates Foundation, government of Botswana
Novartis	Donations of drugs for leprosy	2000 to 2010	All countries affected by leprosy	National health ministries, WHO and NGOs
Novartis	Directly observed therapy for tuberculosis	2003 to 2007	Tanzania	WHO
Novartis	Coartem made available at, or below, cost to malaria patients	2001 and ongoing	All malaria-endemic countries	Chinese for drug development; WHO and Global Fund for delivery
Pfizer	Donations of Diflucan for infections in AIDS patients	2000 and ongoing	Africa, Asia and the Caribbean	WHO, South African government
Sanofi-Aventis	Dengue vaccine project	1993 and ongoing	Thailand	Mahidol University, Bangkok and Thai government
Sanofi-Aventis	Donations of drugs for sleeping sickness	2001 to 2006	Endemic countries in Africa	WHO, MSF, Bill & Melinda Gates Foundation

BOOKS & ARTS

Experimental fiction

Publishers could do a lot more to promote 'lab lit', a genre of novel set in the world of science.

Jennifer Rohn

Art may be said to imitate life, but when it comes to scientists in fiction, the picture is both minimalist and out of focus. Unexpectedly, a recent experiment suggests that publishers' marketing strategies are partly to blame.

Although science fiction thrives, novels containing realistic depictions of scientists plying their trade — a genre I call 'lab lit' — are rare, and those most heralded as literary scientific novels often contain the worst stereotypes of all. A notable example is *Atomised* by Michel Houellebecq, whose scientist embodied not one but all five of the most frequent boffin clichés: arrogance, asexuality, semi-autism, out-of-control experiments and concomitant downfall. Other novels occasionally contain more realistic scientists, but they and their science are rarely central to the plot.

In theory, science should be rich fodder for serious literary fiction if you consider its components: secret knowledge incrementally uncovered, the thrill of investigation, nothing less than the meaning of life and the Universe. The backdrops on offer are fabulous: bustling labs, gleaming machinery, vast telescopes trained on faraway stars, exotic viral epidemics. And the culture is a complex web of urgent human passions and behaviours. Why, then, is this creative treasure so consistently underused, especially in an era when science is immediately relevant to modern life and saturates our media?

One reason is undoubtedly that the adage 'write what you know' hinders the uninitiated, and in recent years a number of well-meaning organizations have tried to get scientists and writers together. These efforts are laudable and should continue, but there is another link in the chain that needs attention: the publishing and marketing phase.

I have spoken to several lab-lit authors who described stern resistance from agents and publishers — some fear that science won't sell, but the main impediment seems to be one of categorization. In a fiercely competitive literary market, genre — knowing what sort of novel one is dealing with — is an important way of helping a book stand out. Groupings such as 'chick lit' and 'historical fiction' made recognizable by characteristic cover designs,



Novel idea: describing fiction set in the laboratory as 'lab lit' could boost book sales.

can function as short cuts for overstimulated consumers.

But what happens when a novel is lab lit? There aren't many similar novels out there, so any that do get written and published are left to fend for themselves in a sea of other well-branded novels. And poor sales will feed back: next time, the publisher might think twice before taking a risk on such a work.

I wanted to know whether lab lit would sell if readers were made aware of it as a genre. To explore this, I performed an experiment with the cooperation of the Gower Street branch of the bookseller Waterstone's in London. The original plan was to display 30 or so lab-lit novels on a table under a promotional poster, but it soon became apparent that we couldn't find enough suitable titles. Even to

achieve the dozen books needed to stock a small upright display, I had to make some compromises. In an ideal world, I wouldn't have needed to venture into the grey areas between pure lab lit and other genres, with Greg Bear's exaggerated *Blood Music* or Patricia Cornwell's forensics-heavy *Unnatural Exposure*, for example. But key books such as Simon Mawer's *Mendel's Dwarf*, Neal Stephenson's *Zodiac* and Carl Djerassi's *Cantor's Dilemma* were deemed corporately "unavailable" (despite their presence on Amazon). After an exhaustive research and repeated liaisons with Waterstone's, I felt lucky to have even a few bona fide specimens, including Jonathan Lethem's *As She Climbed Across the Table*, John McCabe's *Paper* and William Boyd's *Brazzaville Beach*.

The display, initially intended to exist for only a few weeks, was ultimately so successful that it was kept up for five months. And for many of the books displayed, sales increased dramatically, even taking into account the likelihood that displayed books will sell better in general anyway. Of course, we can only draw qualitative conclusions from such a small and uncontrolled experiment. The fact that this particular branch is a major academic bookstore could also have played a role, so it remains to be seen whether we can extrapolate to a wider audience. But the results do suggest that people actually want to read stories about realistic scientists and scientific scenarios — there is a market for such fare, even if this market is not exploited by publishers.

The literary world would do well to recognize this lost opportunity. With millions of people engaged in global scientific research and the daily papers filled with science news, lab lit could be a lucrative new genre. And with public distrust of science as prevalent as ever, it certainly wouldn't hurt science's image to have more lab-lit novels — featuring positively portrayed, realistic scientists — on the best-seller list.

Jennifer Rohn lives in London, UK. She used to do research in molecular biology, and is now editor of LabLit.com magazine.

LABLIT.COM

EXHIBITION

Geological fireworks

Neapolitan artists used to provide pictorial souvenirs for the eighteenth-century grand tourists who trekked up Mount Vesuvius to peer into its smouldering crater. They depicted its eruptions, usually against a night sky to heighten the dramatic effects of glowing, molten lava and trajectories of fiery sparks.

Henry Johnston-Lavis (1856–1914), while working as a doctor in Italy, collected such paintings and prints of historic eruptions and earthquakes, as well as rare books, including William Hamilton's *Campi Phlegraei* (1776–79), an account of his observations on the 'fields of fire' near Naples. Johnston-Lavis also collected albumen prints of photographs, and geological specimens.

After studying both medicine and geology at University College London (UCL), Johnston-Lavis became an authority on the volcanoes of southern Italy. He published the first geological map of Mount Vesuvius in 1891 and was appointed professor of volcanology at the Royal University of Naples in 1893. His observations have aided the reconstruction of past eruptions, and his knowledge of how their nature evolved over time has contributed to the modern study of geological hazards.

An exhibition, *Violent Earth*, drawn from Johnston-Lavis' collection of volcanological material, which he bequeathed to UCL, can be seen at UCL's Strang Print Room on weekday afternoons until 28 April. **C.M.**



UCL ART COLLECTIONS

Living with infection

Diseases and Human Evolution

by Ethne Barnes

University of New Mexico Press: 2005.
484 pp. \$29.95

Tony McMichael

The widespread resurgence of infectious diseases since the 1970s has stimulated many books about this ancient scourge. Following the lightning strike of HIV/AIDS in the early 1980s and its subsequent spreading wildfires, our sensitivity to threats from the realm of infectious disease has been reawakened. The many recent strikes, mostly from viral respiratory diseases, attest to the rising activity of ever-opportunistic microbes in an interconnected and rapidly changing modern world.

This book by the palaeopathologist Ethne Barnes traces the long history of human infectious diseases. Ever since humans first settled in villages, a succession of microbes, mostly animal-derived, have adapted to this auspicious medium. Some have become endemic infections; others make occasional forays from animal sources and may trigger devastating human epidemics. Patterns of infectious disease have changed kaleidoscopically as our forebears' culture evolved from agrarianism to nineteenth-century industrialization. Today we generate ecological niches for microbes through intense food production, greater human mobility, crowded peri-urban poverty and modern medical manoeuvres, such as transfusion and transplantation.

Barnes has digested a voluminous scientific literature and gives an orderly, well-written and comprehensive account of the topic. For a succession of types of infectious disease, she discusses origins and sources, genetic adaptations (of both microbe and human), microbial biology, population-health impacts, clinical features and, in some cases, control policies. The 23 chapters are approximately chronological, encompassing the parasites that first travelled with post-australopithecine hunter-gatherers, the revolution in human-microbial relations ushered in by farming and the consequent rise of various human-adapted infections (malaria, schistosomiasis, trypanosomiasis, tuberculosis, leprosy). Then there's the amplification of infectious diseases by urbanization and, in recent centuries, their spread by seafaring empires. Dramatic epidemics have occurred along the way — Europe, for example, has suffered from the bubonic plague (especially the fourteenth-century Black Death) and syphilis. In the crowded squalor of early industrialization, whole populations, and especially the urban poor, were ravaged by smallpox, cholera, tuberculosis, measles and other infectious diseases. In today's world, influenza is going global; many new infectious diseases are emerging, including HIV/AIDS and severe acute respiratory syndrome (SARS); and surprises have arisen such as Britain's mad cow disease and its human version, variant Creutzfeldt-Jakob disease.

The word 'disease' in the book title is some-

what misleading. Barnes accords little space to non-infectious diseases and, even then, the brief discussions of asthma and other immune disorders, heart disease, diabetes and various cancers tend to highlight the possible contributions of infectious agents. Indeed, the writing here is less enthusiastic and engaged than it is in the author's favoured microbial heartland. Certainly, until early last century the great bulk of (non-violent) deaths everywhere were due to infections and starvation. But today well over half the world's deaths are due to non-infectious diseases. And there is an expanding literature on how the biological legacy of human evolution predisposes us to many of those non-infectious diseases, especially as the living conditions in today's societies deviate ever further from the formative conditions of pre-agrarian life. Maybe there is another book to be written, to round out the story.

The author invokes the discomfiting military idiom that permeates much of the writing about this topic: chapter 2 is titled 'The war between microbes and men'. This language was adopted early in the basic public-health models of infectious disease — in which researchers estimate 'attack rates' and talk about targets, microbial enemies and defence mechanisms. Modern molecular biology has embellished the idiom with notions of molecular missiles, antigenic camouflage and so on. However, this 'us against them' perspective can distort our understanding of the evolutionary basis and ecological complexity of infectious-disease transmission and virulence. In the rapidly changing world we live in today, we need a greater understanding if we are to lessen, proactively, the risks of new infectious diseases arising. Defensive reactivity —

although it is necessary — will not be enough.

There is much interesting detail in this book. The opening chapter sets the stage well, discussing how the intertwined stories of cultural and genetic evolution are fundamental to the emergence and spread of infectious diseases. The book's narrative would have been enriched by a more explicit exploration of these unifying threads throughout. All too briefly, and therefore superficially, the final

several pages consider the high-risk path that we humans are now following. In a "crowded world" that is "out of order", more attention must be paid to how infectious disease may cut swaths through increasingly vulnerable populations in a world of rising microbial mobility.

Tony McMichael is at the National Centre for Epidemiology and Population Health, Australian National University, Canberra, ACT 0200, Australia.

that we distinguish are mediated by culture". They go on to state that "it is impossible to discuss ancient religions and cosmologies in anything but a superficial, periphrastic way without recognizing the input of the human nervous system as it daily produces varied states of consciousness".

The authors define religion broadly to include experience, belief and practice. Without rejecting a position based on Marxist theory, they argue that religion was the driving force behind what Childe referred to as the Neolithic revolution. According to Lewis-Williams and Pearce, "It was religious experience that gave people the power to command the construction of megalithic monuments and to sacrifice animals and very probably human beings in order to keep the cosmos in good order." In other words, animal and human sacrifices "kept the elite in power". The authors urge archaeologists to consider "new types of explanation that do not assume humankind's impotence in the face of environment".

Lewis-Williams and Pearce argue that researchers can use scientific knowledge about consciousness to solve questions in archaeology. They also suggest that mainstream studies of technology and ecological adaptations ignore key variables that drive cultural change, including religion.

Although I accept many of the authors' basic premises, I find it disconcerting that their explanations are not readily refutable. Or, if they are, the authors do not give us clear guidelines on how their hypotheses and interpretations can be tested. This issue also needs to be addressed, particularly with regard to Lewis-Williams' influential work on shamanism.

The authors go further, arguing that their work can be used as a framework in which to analyse current belief in supernatural beings, whether in the form of Christian, Islamic or other kinds of fundamentalism. In this sense they no doubt share my dismay at polls indicating that more than half the population of the United States, including the president, do not accept the validity of evolutionary theory. They comment: "If an American president announces that his decisions are guided by God, alarm bells start ringing." Clearly the authors see their study of Neolithic religion as relevant in the context of today's world.

Like *The Mind in the Cave*, this well produced and finely illustrated book will be of interest to all archaeologists who think that the events of the Stone Age cannot be understood solely by the study of technology, environmental change and calorie counting of the behavioural-ecological school. Most colleagues will not change their research strategies to emulate those presented in *Inside the Neolithic Mind*. But the smart ones will pause a little longer before dismissing the archaeology of religion. ■

Nicholas J. Conard is in the Department of Early Prehistory and Quaternary Ecology, University of Tübingen, Tübingen 72070, Germany.

Unearthing religion

Inside the Neolithic Mind: Consciousness, Cosmos and the Realm of the Gods

by David Lewis-Williams & David Pearce

Thames & Hudson: 2005. 320 pp. £18.95

Nicholas J. Conard

David Lewis-Williams, professor emeritus at the University of the Witwatersrand in South Africa and member of its Rock Art Research Institute, shows no sign of losing the desire to confront his archaeological colleagues with new and controversial ideas.

If his earlier book *The Mind in the Cave* (Thames & Hudson, 2002) rocked the boat of mainstream archaeological science with its innovative and insightful analysis of the origins of art and the replacement of Neanderthals by modern humans, then his new one, *Inside the Neolithic Mind*, co-written with David Pearce, is nothing less than an attempt to capsize the vessel of mainstream archaeology altogether.

The Mind in the Cave contained a fair amount of marxist theory, and opened with a quotation from Karl Marx. The opening chapter of *Inside the Neolithic Mind* also states broad support for marxist approaches and the work to this end by V. G. Childe, but focuses instead on the nature of human consciousness. It presents Coleridge's opium-induced poem

Kubla Khan as an example of a form of consciousness analogous to that experienced by the makers of the earliest monumental architecture at sites including the religious centre of Göbekli Tepe in Turkey and the chamber tombs of Knowth and Newgrange in Ireland.

This discussion of Coleridge's poem is just the start of a literary and scientific tour de force that touches on the works of Dante, Jean-Jacques Rousseau, Aristotle, St Paul, Thomas Aquinas, Rudyard Kipling and other figures who are not the standard fare of scientific archaeology. Many academic and field archaeologists will retrench and find polemic arguments against the authors' unconventional ideas and methods. But I think Lewis-Williams and Pearce have done the scientific community a service by continuing to push the frontiers of archaeological knowledge.

Building on detailed discussions of the neurophysiology and cognitive science of altered and heightened states of consciousness, combined with diverse archaeological and ethnographic evidence, Lewis-Williams and Pearce lay the groundwork for their analysis. Given the multilayered complexity of the book, it is best to turn to the authors' words. Concerning their methods, they argue that their neurological approach "is thus in no way deterministic: all the stages and experiences of consciousness



Can neurological studies shed light on the building of Neolithic religious centres such as Göbekli Tepe?

K. SCHMIDT/ORIENT-ABTEILUNG DAI

CONDENSED-MATTER PHYSICS

Great moments in disorder

Steven T. Bramwell

An array of nanomagnets has been designed to resemble the disordered magnetic state known as 'spin ice'. This could transform our understanding of disordered matter and, potentially, lead to new technologies.

How can we understand disordered states of matter such as liquids, glasses and disordered magnets? First, we must know how they are organized; second, we must know how that organization responds to changes in external constraints such as temperature, pressure or magnetic field. This knowledge is beyond our reach in most cases: disordered states are intrinsically complicated and do not reveal themselves clearly to experiment. But thankfully, there are exceptions. One of them is a magnetic state known as spin ice in which the magnetic moments of ions — their 'spins', in analogy to the property of electron spin — remain disordered even at low temperatures, revealing much about the basic physics of disorder.

On page 303 of this issue, Schiffer and colleagues (Wang *et al.*)¹ report the creation of spin-ice behaviour in an array of nanoscale magnets. Such 'artificial' spin ice, which is stable at room temperature and possesses magnetic moments large enough to be

observed directly, offers a new approach to understanding and exploiting the properties of disordered systems.

In conventional spin ice^{2,3}, magnetic ions form a network of linked tetrahedra (for example, ions of the lanthanide element holmium in the compound holmium titanate). The spins of these ions point either in or out of the tetrahedra (Fig. 1a). The dipolar interaction with their neighbours' magnetic fields favours an in-out arrangement of neighbouring moments, but not all pairs of neighbours can be satisfied simultaneously: the system is 'frustrated'. The best compromise — two moments pointing in, two pointing out, in any one tetrahedron — constitutes the local organizing principle. The same rule controls the hydrogen structure of ice⁴ (Fig. 1b), and is fully satisfied by a huge number of equivalent arrangements of spins (or hydrogen atoms). This means that the chance of achieving an ordered structure is effectively nil. So spin ice, just like normal ice,

remains disordered, with a non-zero entropy (a key measure of disorder) as its temperature approaches absolute zero.

Schiffer and colleagues' artificial spin ice¹ consists of a two-dimensional array of 80,000 elongated magnetic islands, each a few hundred nanometres long. The magnetic moment of every island is aligned parallel to its long axis, as in a bar magnet, and is coupled to its neighbours by the ubiquitous dipolar interaction. For a square geometry, the two in, two out rule is approximately satisfied ('square ice'⁵, Fig. 1c). But because the magnetic moments involved are about three million times bigger than those of holmium ions, they interact more strongly and have less tendency to flip. The artificial spin-ice state is therefore stable at room temperature; for conventional spin ice, this is only the case at temperatures below 1 kelvin. Schiffer and colleagues encouraged their system to settle into a minimum energy state by cycling the applied magnetic field, and

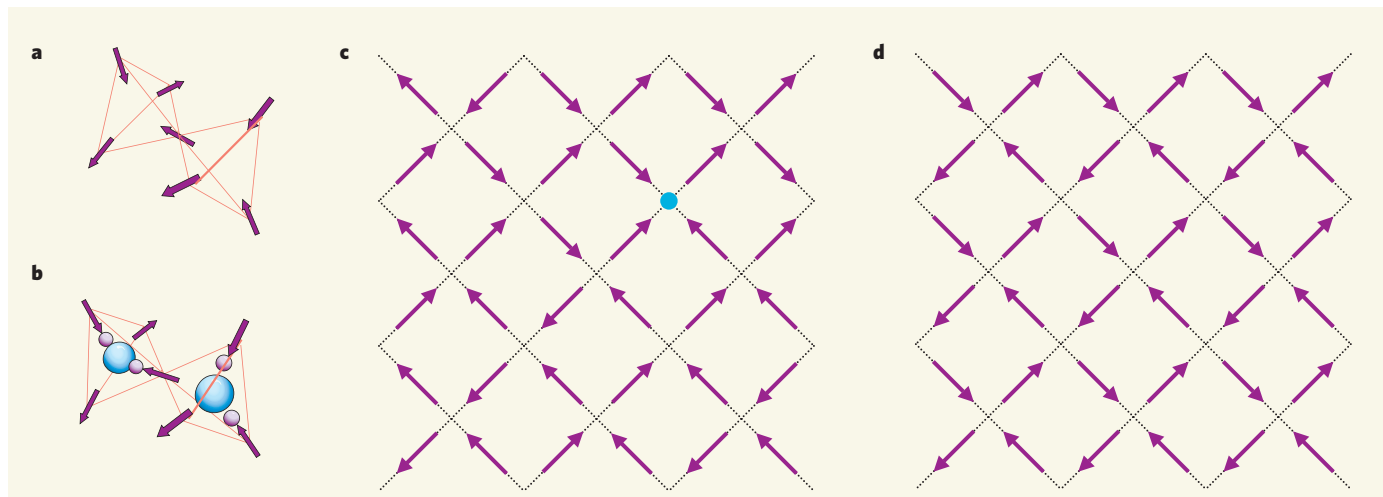


Figure 1 | Organized disorder. **a**, Part of the disordered arrangement of ionic magnetic moments (spins) in spin ice, holmium titanate ($\text{Ho}_2\text{Ti}_2\text{O}_7$): a network of linked tetrahedra with two spins in, two spins out per tetrahedron. **b**, Put a hydrogen atom on the end of every spin of spin ice, and you represent the low-temperature disordered hydrogen structure of (cubic) water ice, H_2O : each hydrogen atom lies on the line connecting two neighbouring oxygen atoms, but is shifted away from the mid-point of that line to ensure that the whole structure is a hydrogen-bonded network of v-shaped water molecules (oxygen atoms, large circles; hydrogen atoms,

small circles). **c**, Here arrows represent the magnetic moments of nanoscale magnetic islands in the disordered state of the artificial 'square' spin ice created by Wang *et al.*¹. Each corner has two arrows pointing in, and two pointing out, but within this constraint the direction of the arrows on any one corner is arbitrary: the two in, two out condition is not sufficient to define an ordered state. The blue circle marks a defect where the two in, two out configuration is not maintained. **d**, An ordered pattern in which the arrows of alternate squares have the same configuration. This is possibly the true minimum energy state of artificial spin ice.

then determined the directions of individual magnetic moments using a technique known as magnetic force microscopy. Statistical analysis of these directions confirmed that a spin-ice state, characterized by a preponderance of the two in, two out configuration, had indeed been created.

Conventional spin ice lends itself well to experimentation. Its disordered spin arrangements have been imaged by neutron scattering, and applying magnetic fields to it has revealed disordered phases that respond to neutron scattering and bulk measurements as if they consisted of stacks of independent chains or sheets — so as if they have only one or two dimensions, rather than three. Phase transitions that, like the liquid–liquid or liquid–gas transitions in more complex matter, involve no change in structural symmetry have also been observed⁶. In such ‘symmetry-sustaining’ transitions, what is ordered remains ordered, and what is disordered remains disordered.

But experiments on conventional spin ice raise interesting questions. First, what is the true origin of the two in, two out rule? The dipolar interaction between magnetic moments is long-range, so any explanation that considers only near neighbours must be incomplete. This question has largely been answered: remarkably, the standard model of spin ice predicts that the two in, two out rule emerges from the many-bodied dipolar interaction of some 10^{23} magnetic moments^{7,8}. Second, is there a ‘true’ ordered ground state (Fig. 1d)? The question is also pertinent to normal ice⁹, as such a minimum-energy state is required by the third law of thermodynamics, which states that entropy approaches zero as temperature tends to absolute zero. An ordered ground state is also predicted by the standard model of spin ice, but is not observed experimentally. Finally, what microscopic factors drive phase transitions in an applied magnetic field? This question, too, remains essentially open.

Artificial spin ice could, in principle, help to supply further insight into these problems. It could, for example, be engineered to have different, controllable interactions and then be subjected to temperature changes or magnetic fields to mimic the behaviour of conventional spin ice. Direct imaging with magnetic force microscopy could be used to identify and understand individual defects in the spin-ice state. Such defects (Fig. 1c) cannot be imaged by neutron scattering on conventional spin ice, but might be crucial in determining its properties.

Bringing spin ice to room temperature could also inspire technological applications. In magnetic-memory media, information is encoded into the magnetic moments of ferromagnetic grains. The drive to increase the density of memory bits in such media will mean smaller, more strongly magnetized elements that are more closely spaced¹⁰. This trend will amplify the dipolar interaction and its consequences¹¹. Experiments with spin ice,

however, show how to create a dense array of magnetic elements, which, although they interact, retain many states in which information could potentially be encoded.

This is for the future. As a replica of conventional spin ice, artificial spin ice is not perfect: the two in, two out configuration is maintained only approximately, and the system should actually prefer an alternative, ordered state¹² (Fig. 1d). Its failure to find this state might reflect the inefficiency of the energy-minimization protocol involving magnetic field cycling. Despite its limitations, however, Schiffer and colleagues’ invention¹ does emphasize the potential of designed magnetic arrays, not only as model systems for the study of disorder, but also as the basis of technological applications. ■

Steven T. Bramwell is in the Department of Chemistry, University College London,

20 Gordon Street, London WC1H 0AJ, UK, and the London Centre for Nanotechnology, 2–16 Torrington Place, London WC1E 7HN, UK. e-mail: s.t.bramwell@ucl.ac.uk

1. Wang, R. F. *et al.* *Nature* **439**, 303–306 (2006).
2. Harris, M. J., Bramwell, S. T., McMorrow, D. F., Zeiske, T. & Godfrey, K. W. *Phys. Rev. Lett.* **79**, 2554–2557 (1997).
3. Ramirez, A. P., Hayashi, A., Cava, R. J., Siddharthan, R. & Shastri, B. S. *Nature* **399**, 333–355 (1999).
4. Pauling, L. J. *Am. Chem. Soc.* **57**, 2680–2684 (1935).
5. Lieb, E. H. *Phys. Rev. Lett.* **18**, 692–694 (1967).
6. Sakakibara, T., Tayama, T., Hiroi, Z., Matsuhira, K. & Takagi, S. *Phys. Rev. Lett.* **90**, 207205 (2003).
7. Melko, R. G. & Gingras, M. J. P. *J. Phys. Cond. Mat.* **16**, R1277–R1319 (2004).
8. Isakov, S. V., Moessner, R. & Sondhi, S. L. *Phys. Rev. Lett.* **95**, 217201 (2005).
9. Singer, S. J. *et al.* *Phys. Rev. Lett.* **94**, 135701 (2005).
10. Moser, A. *et al.* *J. Phys. D* **35**, R157–R167 (2002).
11. Martin, J. I. *et al.* *J. Magn. Magn. Mater.* **256**, 449–501 (2003).
12. De’Bell, K., MacIsaac, A. B., Booth I. N. & Whitehead, J. P. *Phys. Rev. B* **55**, 15108–15118 (1997).

CANCER BIOLOGY

Signatures guide drug choice

Julian Downward

Cancer drugs are increasingly designed to target specific cell-signalling pathways. When, and in what combination, these drugs should be used might be judged by analysing the gene expression signature of the tumour.

Current approaches to the design of drugs against cancer assume that almost all tumours escape normal growth regulation by usurping a few of the dozen or so key cell-signalling pathways. However, pathways can be activated at different points, so it is not always easy to tell which signalling mechanism has been activated by looking for mutations in known cancer-associated genes (oncogenes, or tumour-suppressor genes). If the gene at the top of a signalling cascade is unaffected, for instance, one cannot assume that the pathway is not involved, as a factor further downstream might have been activated. It can thus be hard to predict the best treatment for a particular tumour. Two papers in this issue^{1,2} address this problem in different ways, and provide a potential strategy for choosing the most effective combination of therapies based on the gene expression signature of a tumour.

Tumour cells seem to rely heavily on the continued activation of one or two pathways — a phenomenon termed oncogene addiction — whereas normal cells use a broader range of signals³. This, combined with the damage accrued through the reckless lifestyle of the cancer cell, provides an Achilles’ heel that might be exploited therapeutically by targeting pathways activated by oncogenes such as RAS, SRC and MYC. The RAS pathway (Fig. 1), for example, can be activated in many different ways in tumours, including mutation of the RAS oncogene itself (seen in 40% of lung

cancers) and of the BRAF oncogene, the next factor in the pathway (mutated in some 60% of melanomas)⁴. Thus, looking for evidence of an initiating mutation can make it hard to identify all tumours in which this pathway has been activated.

Equally, looking for a single downstream indicator of pathway activation, such as phosphorylation of the enzyme ERK (one of the final steps in the RAS pathway), can also be problematic. Negative feedback loops, such as the induction of phosphate-removing enzymes that target ERK, can attenuate the steady-state phosphorylation of ERK. In addition, branching of the pathway can mean that other targets might be more important in certain circumstances (Fig. 1).

One step in the RAS pathway that is being targeted by candidate drugs is MEK, an enzyme that is directly activated by BRAF, and that is thought to be responsible for much of the downstream signalling from RAS (Fig. 1). To see whether MEK inhibitors could be useful for treating all tumours with aberrant RAS signalling, Solit *et al.* (page 358)¹ tested human tumour cell lines carrying mutations in BRAF or RAS for sensitivity to these drugs. Cells bearing an activating BRAF mutation were extremely sensitive to MEK inhibitors, both *in vitro* and when transplanted into immunodeficient mice. By contrast, cells with an activating RAS mutation showed much lower and more variable sensitivity to these inhibitors.

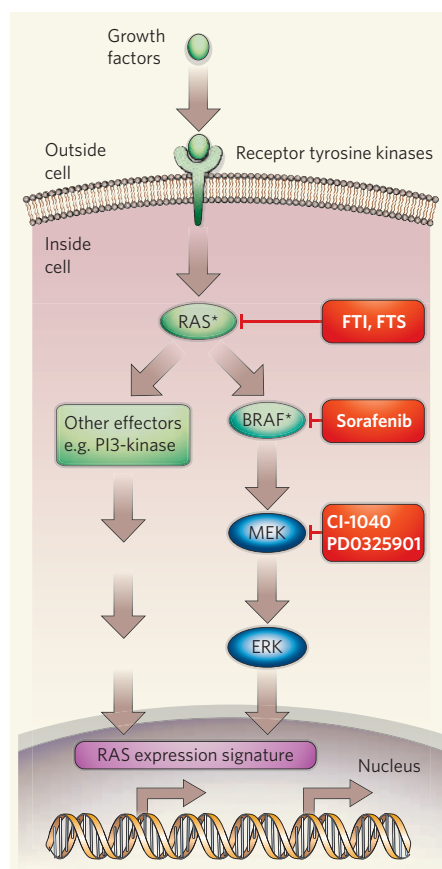


Figure 1 | Oncogene activation, transcription signatures and drug sensitivity. Growth factors activate RAS through receptor tyrosine kinases, leading to stimulation of the BRAF, MEK and ERK cascade, but also of other pathways, including that involving phosphatidylinositol (PI) 3-kinase. Activation of the RAS pathway in tumours can occur through mutation or overexpression of the components shown in green. Bild *et al.*² define unique gene expression signatures for five oncogenic pathways, including that for RAS. These can be used to assess the level of activity of each pathway in a cancer cell and the sensitivity to drugs targeting those pathways (red). Solit *et al.*¹ show that activating mutations (*) in BRAF make the cells sensitive to inhibitors of MEK, whereas those in RAS do not, presumably because it can signal through an alternative pathway.

The fact that the mutational status of BRAF predicts sensitivity to inhibition of MEK suggests that all oncogenic signalling from BRAF is mediated through the activity of its direct target MEK. However, signalling from RAS bifurcates to several downstream targets in addition to BRAF and MEK, including key pathways such as that involving phosphatidylinositol 3-kinase. Thus, inhibition of MEK may not be sufficient to inhibit cell proliferation triggered by these pathways, at least in some situations⁵. The clear implication is that the MEK inhibitors being developed as potential drugs⁶ should be tested on tumours bearing activating BRAF mutations, such as melanomas, and not on tumours with activating RAS mutations, such as pancreatic and lung carcinomas. Indeed, the mutational

status of BRAF should be used to stratify patients in any such trials.

What, then, of situations in which pathway activation might not be caused by a single oncogenic mutation? Bild *et al.* (page 353)² used microarrays to analyse the gene expression profiles of human mammary epithelial cells in which five key oncogenic pathways had been activated — by mutational activation of the MYC, RAS, SRC or β -catenin proteins, or by loss of the Rb tumour-suppressor gene. In each case, the authors defined a signature of a hundred or so genes whose expression correlated with activation of the specific pathway.

Similar individual signatures have been characterized before^{7–10}, but here Bild *et al.* used them simultaneously to analyse the activation state of each of the pathways in a range of human and mouse tumours. The signatures successfully predicted the activating mutation in several mouse models of cancer and in human lung cancers bearing RAS mutations. In addition, predictions of the degree of deregulation of each pathway could be used as a basis for categorizing tumours into clusters that showed marked correlations with clinical outcome. For example, in lung cancer, deregulation of MYC, RAS, SRC and β -catenin together correlated with particularly poor patient survival.

The signatures for RAS and SRC pathway activation accurately predicted the *in vitro* sensitivity of a broad range of human tumour cell lines to drugs targeting the mutationally activated versions of these proteins. However, the RAS signature did not correlate with the levels of activated RAS protein in the cell, presumably because the same signature of gene

expression can be achieved by activation of upstream or downstream components of the pathway. One might conclude that analysis of gene expression signatures is a more reliable predictor of pathway activation across different tumour types than analysing the mutation state or expression level of a given oncogene. This would be particularly true for complex branching pathways such as the RAS one.

So from a single microarray (which can analyse the expression levels of all the genes involved at once) it should be possible to determine the extent to which different signalling pathways are activated and what combination of pathway-specific drugs might, where available, be most effective. In favourable cases, such as those in which BRAF is mutated, the state of a single oncogene may be sufficient to predict response to a single drug, although this will probably be the exception rather than the rule. The foundations may have been laid for the development of truly rational combination therapies for multigene cancers.

Julian Downward is at the Cancer Research UK London Research Institute, 44 Lincoln's Inn Fields, London WC2A 3PX, UK.
e-mail: downward@cancer.org.uk

1. Solit, D. B. *et al.* *Nature* **439**, 358–362 (2006).
2. Bild, A. H. *et al.* *Nature* **439**, 353–357 (2006).
3. Weinstein, I. B. *Science* **297**, 63–64 (2002).
4. Downward, J. *Nature Rev. Cancer* **3**, 11–22 (2003).
5. Mao, J. H. *et al.* *Genes Dev.* **18**, 1800–1805 (2004).
6. Sebolt-Leopold, J. S. & Herrera, R. *Nature Rev. Cancer* **4**, 937–947 (2004).
7. Neiman, P. E. *et al.* *Proc. Natl Acad. Sci. USA* **98**, 6378–6383 (2001).
8. Huang, E. *et al.* *Nature Genet.* **34**, 226–230 (2003).
9. Pavay, S. *et al.* *Oncogene* **23**, 4060–4067 (2004).
10. Sweet-Cordero, A. *et al.* *Nature Genet.* **37**, 48–55 (2005).

ATMOSPHERIC CHEMISTRY

Biogenic bromine

Ross J. Salawitch

Among other effects, bromine released by biological processes in the oceans apparently reduces ozone levels in the troposphere. This source may be a link between atmospheric composition and climate change.

Bromine compounds from organic halogens (halons) used in fire extinguishers and from methyl bromide, which has anthropogenic and natural sources, cause about half of the chemical loss that results in the Antarctic 'ozone hole' in the stratosphere. Low levels of ozone in the atmosphere's lowermost layer, the troposphere, during the polar spring result from bromine released from melting sea ice and 'frost flowers'. But it is also becoming evident that bromine produced by natural processes in the ocean can influence the composition of both the troposphere and the stratosphere. Much research is being devoted to understanding the sources and sinks of

these organic bromine compounds, and their effects in the atmosphere.

Yang *et al.*¹, for example, writing in the *Journal of Geophysical Research*, present simulations of how the troposphere is affected by the inorganic bromine molecules that are released when biogenic organic bromine decomposes. They calculate that levels of tropospheric ozone are reduced by 5–30%, depending on location and season, relative to a simulation that does not take bromine into account. Two mechanisms are involved. One is the direct catalytic loss of ozone in a reaction involving bromine monoxide (BrO). The other is reduced production of ozone; this is caused

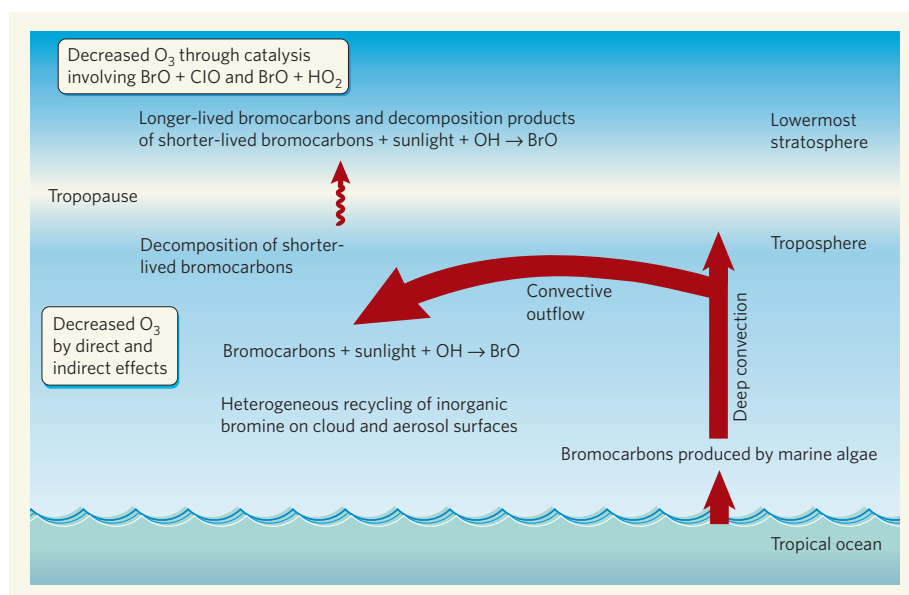


Figure 1 | Ocean, atmosphere and bromine. Bromocarbons produced by algae in the tropical oceans¹² are lofted to the tropopause by deep convection in the atmosphere. Shorter-lived bromocarbons, such as bromoform (CHBr_3), decompose in the troposphere, and convective outflow provides a steady supply of bromine to the global troposphere. Heterogeneous reactions of inorganic bromine on clouds and aerosols¹³ may further increase tropospheric BrO. Reduced levels of ozone occur through a direct effect^{1,2} (catalytic loss initiated by the reaction of BrO with HO_2) and an indirect effect (reduced production owing to decreased NO_x , caused by the hydrolysis of bromine nitrate^{1,3}). Longer-lived bromocarbons, such as dibromomethane (CH_2Br_2) and bromochloromethane (CH_2BrCl), and the decomposition products of the shorter-lived species, may slowly ascend across the tropopause. In the lowermost stratosphere, BrO increases ozone depletion by anthropogenic chlorofluorocarbons through catalytic loss initiated by the reaction of BrO with ClO; bromine is also involved in photochemical loss of ozone through the $\text{BrO} + \text{HO}_2$ cycle⁶.

by decreased levels of nitrogen oxides, and is triggered by the reaction of another inorganic bromine species on the surface of clouds and aerosols. Other modelling work^{2,3} has reported similar reductions in ozone and nitrogen oxides by oceanic bromocarbons.

Yang *et al.*¹ show that BrO levels in their model range from 1 to 8 parts per trillion (p.p.t.). Their calculations are consistent with observations from the Global Ozone Monitoring Experiment (GOME) that indicate much larger abundances of total-atmospheric BrO than can be accounted for by standard models^{4,5}. These models, which assume that atmospheric bromine originates from halons and methyl bromide alone, produce abundances of BrO that are a factor of two to three less than the GOME observations^{5,6}. Because of their relatively long lifetimes in the atmosphere, halons and methyl bromide decompose mainly in the stratosphere. Furthermore, standard models assume that inorganic bromine from halons and methyl bromide is rapidly and efficiently washed out of the troposphere, resulting in near-zero levels of BrO there. Yang *et al.* show that levels of tropospheric BrO might be high because decomposition of the shorter-lived, oceanic biogenic bromine compounds occurs primarily in the troposphere.

One of the puzzles for atmospheric chemists is precisely where in the atmosphere this 'excess bromine' revealed by GOME resides. Yang *et al.*¹ show that balloon-borne measurements

of tropospheric BrO profiles⁷, ranging from 0.5 to 2.0 p.p.t., are broadly consistent with their model calculations. But ground-based measurements present an uncertain picture, suggesting variously that all⁵ or almost none⁸ of the 'excess BrO' resides in the troposphere. Balloon and aircraft measurements indicate that much of it may reside in the lowermost stratosphere, with the source being the longer-lived oceanic organic bromine species as well as the decomposition products of shorter-lived species^{6,9}. Satellite observations of BrO profiles obtained by the Scanning Imaging Absorption Spectrometer for Atmospheric Chartography instrument suggest contributions to 'excess BrO' from both the troposphere and stratosphere¹⁰. This BrO in the lowermost stratosphere results in greater depletion of stratospheric ozone at mid-latitudes, particularly during times of increased aerosol levels from volcanic activity. This is because BrO provides an efficient reaction partner for the chlorine monoxide (ClO) derived from anthropogenic compounds such as chlorofluorocarbons⁶.

Researchers are also investigating where in the oceans bromocarbon production occurs. One source is known to be marine algae in coastal regions¹¹. But another is evidently the open, tropical oceans¹². This could be of particular importance because the depth of convection in this region of the atmosphere can quickly loft bromocarbons to the tropopause

(the boundary between the troposphere and stratosphere) — so, perhaps, explaining both the ubiquitous nature of BrO throughout the troposphere^{1,2} and the stratospheric supply of bromine produced from these relatively short-lived compounds¹².

Understanding the photochemistry of oceanic bromocarbons is essential for assessing the fate of these compounds in the atmosphere. Most importantly, the heterogeneous chemistry — the reactions of gas-phase molecules on aerosol and cloud surfaces — is being quantified. For example, it has been shown that labile bromine compounds are released to the gas phase by reactions involving soluble, inorganic species of bromine on solutions of sulphuric acid¹³. Heterogeneous reactions may result in considerably less efficient 'aerosol wash-out' of inorganic bromine than is assumed in many studies, including that of Yang *et al.*¹. But much work remains to define the aerosol-related photochemistry of bromine species for a range of atmospheric conditions.

Several investigations^{1–3,6} suggest that models may need to include the influence of biogenic bromine to represent accurately the photochemistry affecting ozone in both the troposphere and stratosphere (Fig. 1). As noted above, Yang *et al.*¹ show that oceanic bromocarbons may reduce present-day levels of ozone and nitrogen oxides throughout the troposphere. Tropospheric ozone is both an air pollutant and a greenhouse gas. Accurate modelling of the nitrogen-oxide budget resulting from natural processes, including lightning, is a prerequisite for assessing perturbations to ozone caused by the industrial release of nitrogen oxides.

Finally, Yang *et al.*¹ and others^{11,12} speculate that oceanic production of bromocarbons may be a new link between climate change and the composition of the global atmosphere. Observations indicate a link between wind-driven oceanic upwelling, production of bromocarbons and transport of these compounds to the tropical tropopause¹². Oceanic production of bromocarbons is sensitive to temperature, nutrient supply and upwelling^{11,12}. If in the future the atmosphere becomes warmer and more strongly convective, increased BrO might result in lower levels of ozone in both the troposphere and stratosphere. Finding out whether biogenic bromine does indeed link ocean biology, climate and atmospheric composition will require concerted studies from those involved in laboratory measurement, field observation and modelling. ■

Ross J. Salawitch is at the Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California 91109, USA.
e-mail: ross.salawitch@jpl.nasa.gov

1. Yang, X. *et al.* *J. Geophys. Res.* **110**, D23311; doi:10.1029/2005JD006244 (2005).
2. von Glasow, R. *et al.* *Atmos. Chem. Phys.* **4**, 2481–2497 (2004).
3. Lary, D. *Atmos. Chem. Phys.* **5**, 227–237 (2005).

4. Chance, K. *Geophys. Res. Lett.* **25**, 3335–3338 (1998).
5. van Roozendaal, M. *et al. Adv. Space Res.* **29**, 1661–1666 (2002).
6. Salawitch, R. J. *et al. Geophys. Res. Lett.* **32**, L05811; doi:10.1029/2004GL021504 (2005).
7. Fitzenberger, R. *et al. Geophys. Res. Lett.* **27**, 2921–2924 (2000).
8. Schofield, R. *et al. J. Geophys. Res.* **109**, D14304; doi:10.1029/2003JD004463 (2004).
9. Pfeilsticker, K. *et al. Geophys. Res. Lett.* **27**, 3305–3308 (2000).
10. Sinnhuber, B.-M. *et al. Geophys. Res. Lett.* **32**, L20810; doi:10.1029/2005GL023839 (2005).
11. Carpenter, L. J. & Liss, P. S. *J. Geophys. Res.* **105**, 20539–20547 (2000).
12. Quack, B. *et al. Geophys. Res. Lett.* **32**, L23505; doi:10.1029/2004GL020597 (2004).
13. Iraci, L. T. *et al. Atmos. Chem. Phys.* **5**, 1577–1587 (2005).

PLANT BIOLOGY

Absciscic acid in bloom

Julian I. Schroeder and Josef M. Kuhn

To survive environmental stresses, plants must respond to the hormone absciscic acid. The receptors for this hormone have remained elusive, but one receptor with unique functions in flowering has now been identified.

When plants experience drought or cold, they cannot get themselves a glass of water or move to a warmer place. Instead, their ability to survive lack of water, extreme temperatures and such stresses as high salt levels relies heavily on a plant hormone called absciscic acid (ABA). Despite their importance, the genes that encode the cellular receptors for this hormone have not been identified. On page 290 of this issue, however, Razem *et al.*¹ describe the characterization of a protein that binds to messenger RNA and that also binds ABA and controls ABA-dependent flowering in the model plant *Arabidopsis*.

The question of how plants cope with the recurring stresses of drought, cold and salinity not only engages plant scientists, agronomists, ecologists and climatologists. It also increasingly demands the attention of politicians, given that in arid regions across the globe more than 80% of the available fresh water is consumed by agriculture². Many avenues of research have shown that ABA is a key player in such stress resistance. Responses mediated by this hormone lead to the induction of complex tolerance mechanisms to drought, cold, salinity and wounding, including the control of closure of the stomatal pores in leaves to reduce water loss³.

Genetic screens with various twists have elucidated ABA signal-transduction mechanisms that act downstream of ABA sensing³. But genes that encode ABA-binding receptor proteins have remained unidentified. This might be because plant genomes have large numbers of homologous — closely related — genes that probably have overlapping functions, or because an ABA receptor is essential, such that plants with mutations in the receptor gene would not survive. Research on ABA signalling is also revealing the robustness of an intricate signal-transduction network. This can limit traditional 'forward' genetic approaches⁴, because a mutation in one pathway may be side-stepped to a degree by using another route that transmits the signal.

Razem and colleagues¹ have used an alternative, biochemical approach. They isolated a barley protein that has ABA-binding activity, named ABAP1 (ref. 5), and investigated whether a homologue of ABAP1 functions in an ABA response in *Arabidopsis*. Their work shows that an RNA-binding protein called FCA binds to ABA and is regulated by it, and that FCA is involved in a less well-studied function of ABA — the inhibition of flowering.

The *Arabidopsis* FCA protein is homologous to the barley ABAP1 protein in its carboxy-terminal half and, like ABAP1, it has a high affinity for active ABA analogues^{1,5}. (Its dissociation constant, K_d , is 19 nM.) The FCA protein is a component of the so-called autonomous flowering pathway, which reduces the activity of the flowering repressor *Flowering Locus C*, or *FLC* (Fig. 1, overleaf)^{6,7}. Two structural regions of the FCA protein are of particular relevance: a protein-interaction region known as a WW domain, and an RNA-recognition motif^{6,7}. The WW domain allows FCA to interact with the protein FY, which is an mRNA processing factor⁶.

The FCA–FY complex negatively regulates expression of the flowering repressor *FLC*. It also reduces the amount of functional FCA protein through a negative feedback loop by adding a premature polyadenylation tail to a truncated form of the FCA precursor mRNA^{6,7}. In this negative feedback loop, polyadenylation of the truncated FCA precursor mRNA results in a shortened mRNA, and thus in non-functional FCA protein (Fig. 1). Several reports have established a link between RNA-processing proteins and ABA signalling^{8–12}. But we don't yet know whether these mRNA-processing proteins, which affect ABA action, are components of an FCA-like ABA stress-response pathway.

In the new work, Razem *et al.*¹ report that the FCA–FY complex dissociates when ABA binds to FCA, making the complex non-functional (Fig. 1). As a result, premature polyadenylation of the truncated FCA precursor



50 YEARS AGO

"Benjamin Franklin's Purse"

— In connexion with the 250th anniversary of the birth of Benjamin Franklin (1706–90), the Department of Mineralogy of the British Museum (Natural History) is exhibiting the asbestos purse sold by Franklin to Sir Hans Sloane in 1725. The purse came into the collections of the British Museum at the time of its foundation in 1753, on the death of Sir Hans Sloane. It had lain unrecognized for many years and was identified in 1938. On his arrival in Britain, Franklin worked for some time as a compositor and sought to augment his income by the sale of some 'curiosities' which he had brought with him from America. One of these, the asbestos purse, is referred to by Franklin in his autobiography... The purse represents perhaps the earliest specimen of asbestos from North America to reach Great Britain. From *Nature* 21 January 1956.

100 YEARS AGO

The study of a few of our British stone monuments from an astronomical point of view [may give information on] the order of succession of the various swarms of immigrants who set out the various systems of alignments... I have evidence that the risings of stars, as well as of the sun, were observed in some of the circles... some circles used in the worship of the May year were in operation in 2000 B.C., and there was a change of cult about 1600 B.C., or shortly afterwards, in southern Britain, so definite that the changes in the chief orientation lines in the stone circles can be traced. To the worship of the sun in May, August, November, and February was added a solstitial worship in June and December. The easiest explanation is the advent of a new swarm of immigrants about that date. The associated phenomena are that the May–November Balder and Beltaine people made much of the rowan and maythorn. The June–December people brought the worship of the mistletoe. From *Nature* 18 January 1906.

50 & 100 YEARS AGO

mRNA is abolished. Thus, ABA causes accumulation of full-length *FCA* mRNA. Razem *et al.* show that ABA causes a dramatic increase in *FLC* mRNA, which in turn would delay the transition to flowering. Consistent with this model, the authors report that ABA causes a delay in flowering in *Arabidopsis*. As *Arabidopsis* plants can flower early in response to drought, which increases ABA production, the ABA–*FCA* response may be overridden during this response¹³. Possible modulation mechanisms during drought stress could be investigated by analysing the newly revealed direct ABA regulation of *FCA* mRNA (full-length versus truncated) and the strong ABA-induced increase in levels of *FLC* mRNA.

Interestingly, the RNA-recognition motif in *FCA* is absent in the barley *ABAP1* protein⁵. Indeed, ABA-binding studies of *Arabidopsis FCA* in which the protein lacked specific structural regions show that ABA-binding activity lies in the carboxy-terminal half of *FCA*, which does share homology with *ABAP1* (ref. 1).

Razem *et al.*¹ went on to show that in plants with a loss-of-function mutation in *FCA*, the ABA-induced closing of stomatal pores and inhibition of seed germination — two classical ABA responses — were not impaired. Furthermore, ABA inhibition of flowering was not affected in two dominant ABA-insensitive mutants, *abi1-1* and *abi2-1*, in which most of the stress-related ABA responses are impaired. Thus, other ABA receptors are needed to explain the classical ABA signalling responses to stress. The hunt could be on to characterize

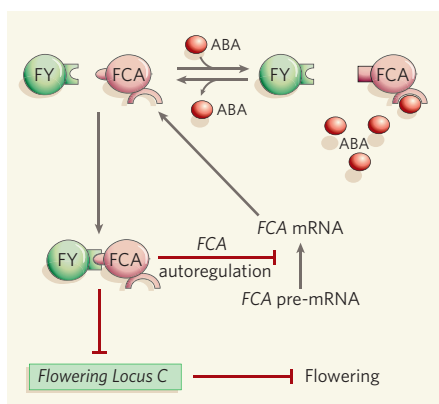


Figure 1 | Absciscic acid, RNA metabolism and control of flowering in plants. Binding of two proteins, *FCA* and *FY*, to one another results in a decrease in expression levels of *Flowering Locus C* (*FLC*), causing a transition from vegetative growth to flowering. The *FCA*–*FY* complex also causes synthesis of a truncated, non-functional *FCA* messenger RNA in a negative feedback loop that results in fewer full-length *FCA* mRNA transcripts and less *FCA* protein^{6,7}. Razem *et al.*¹ report that binding of abscisic acid (ABA) to *FCA* abolishes the interaction of *FCA* with *FY*, leading to an increase in full-length *FCA* transcripts and — through increased *FLC* activity — a delay in flowering. Red lines depict negative regulation. (Diagram modified from a figure provided by R. Hill.)

homologues to the ABA-binding carboxy terminus of *FCA*¹ and barley *ABAP1*. A simple search of protein databases reveals only one distant *FCA* homologue in the *Arabidopsis* genome. Alternatively, the *FCA* and *ABAP1* proteins provide an opportunity to elucidate

the structure of an ABA-binding pocket, which may reveal important sub-domains and structural constraints for ABA binding.

A door to understanding ABA perception has been opened. The binding of ABA to *FCA* and *ABAP1* is apparently a further example of newly emerging mechanisms by which plant growth regulators mediate their responses. Further questions arise with each advance. Plant scientists will need to keep on trekking to illuminate how their immobile lab subjects perceive abscisic acid when faced with drought, cold and salinity.

Julian I. Schroeder and Josef M. Kuhn are at the Division of Biological Sciences, Cell and Developmental Biology Section, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0116, USA. e-mail: julian@biomail.ucsd.edu

1. Razem, F. A., El-Kereamy, A., Abrams, S. R. & Hill, R. D. *Nature* **439**, 290–294 (2006).
2. Delmer, D. P. *Proc. Natl Acad. Sci. USA* **102**, 15739–15746 (2005).
3. Finkelstein, R. R., Gampala, S. S. & Rock, C. D. *Plant Cell* **14**, S15–S45 (2002).
4. Hetherington, A. M. & Woodward, F. I. *Nature* **424**, 901–908 (2003).
5. Razem, F. A., Luo, M., Liu, J.-H., Abrams, S. R. & Hill, R. D. *J. Biol. Chem.* **279**, 9922–9929 (2004).
6. Simpson, G. G., Dijkwel, P. P., Quesada, V., Henderson, I. & Dean, C. *Cell* **113**, 777–787 (2003).
7. Quesada, V., Macknight, R., Dean, C. & Simpson, G. G. *EMBO J.* **22**, 3142–3152 (2003).
8. Lu, C. & Fedoroff, N. *Plant Cell* **12**, 2351–2366 (2000).
9. Hugouvieux, V., Kwak, J. M. & Schroeder, J. I. *Cell* **106**, 477–487 (2001).
10. Xiong, L. *et al. Dev. Cell* **1**, 771–781 (2001).
11. Li, J. *et al. Nature* **418**, 793–797 (2002).
12. Nishimura, N. *et al. Plant J.* **44**, 972–984 (2005).
13. Levy, Y. L. & Dean, C. *Plant Cell* **10**, 1973–1989 (1998).

CHEMICAL ECOLOGY

In defence of maize

They can't run and they can't hide.

But with those reactions denied to them, plants have evolved an extensive and varied repertoire for responding to threats to their well-being. Apart from the tolerance mechanisms discussed above by Schroeder and Kuhn, such responses can include calling upon insect allies to deal with pests that would otherwise damage or destroy the plant by eating it.

Christiane Schnee and colleagues have started to dissect one such signal system — that used by maize seedlings when they are attacked by caterpillars (*Proc. Natl Acad. Sci. USA* doi:10.1073/pnas.0508027103; 2005). The plant signals consist of volatile chemicals, which in the case investigated by Schnee *et al.* attract female wasps of the species *Cotesia marginiventris*. The wasps lay their eggs

in the caterpillars, with predictably unhappy results for the latter.

Maize emits a cocktail of volatile defence signals in response to an attack by herbivores, and the general difficulty in studying the process is identifying which constituent of these complex blends has which effect. From investigations of maize biosynthetic pathways, Schnee *et al.* first isolated an enzyme, a terpene synthase dubbed TPS10, that is responsible for producing most of the herbivore-induced volatiles.

But that was only an initial step. To look into the biological effects of these terpenes, the authors used genetic engineering to insert the gene that encodes TPS10 into *Arabidopsis*, the standard lab plant for biologists. The transgenic *Arabidopsis* plants were then used



in experiments in which female *C. marginiventris* could choose between the options offered in an 'olfactometer' (fresh air and untransformed *Arabidopsis* being the other choices).

The main result to emerge was that the wasps indeed showed a strong preference for the plant that produced the TPS-mediated terpene — but only after they had learned the association between the defence signal and the host by having previously

laid eggs in the caterpillar host.

This approach, say Schnee *et al.*, is an example of the value of using transgenic technology to study the effects of complex volatile compounds. Even when the compounds themselves or their constituents are not available, if the genes concerned have been identified, the ecological influences of these signals can nonetheless be investigated in genetically transformed plants.

Tim Lincoln

NUCLEAR PHYSICS

Odd couple decays

Juha Äystö

The decay of proton-rich nuclei by the emission of a single proton has been known about for some time, and is well understood. The latest observation of two-proton emission, however, will provoke some head-scratching.

On page 298 of this issue¹, Mukha and colleagues report the simultaneous emission of two protons from a complex, long-lived state of the silver isotope ^{94}Ag , which has an odd number of protons. This type of radioactive decay is expected only for proton-rich nuclei with an even number of protons — so the observation leaves nuclear physicists with some explaining to do.

Whether an atomic nucleus is stable or decays depends on the interplay between two fundamental forces: the short-range, attractive strong nuclear force and the longer-range, repulsive electromagnetic Coulomb force. Whereas the strong force acts between all the nucleons (protons and neutrons) that make up the nucleus, the Coulomb force acts only between protons. This means that in nuclei that are extremely rich in protons — said to lie beyond the 'proton drip-line' (Fig. 1) — the strong force can no longer bind all protons, and such nuclei can decay through the emission of one or two protons.

In these rare radioactive decays, protons tunnel quantum mechanically out of the nucleus, through the energy barrier formed by the combined effect of the strong and Coulomb forces. In the lowest allowed energy state of a nucleus, it is energetically favourable for protons or neutrons to pair up. Single-proton emission is therefore expected to occur among proton-rich nuclei that have an odd number of protons, whereas two-proton emission should be characteristic of nuclei with an even number of protons.

Single-proton radioactivity was discovered² several decades ago in the decay of an excited state of cobalt-53 to the ground state of iron-52. Today, about 30 different single-proton radioactivities are known³ for nuclei with proton numbers between 50 and 84, and the phenomenon is fairly well understood theoretically. In contrast, two-proton radioactivity was observed only recently^{4,5} — from the ground states of isotopes of iron, nickel and zinc, ^{45}Fe , ^{48}Ni and ^{54}Zn , all of which have an even number of protons — and information on the energy and angular correlations of the emitted protons is incomplete. The emission

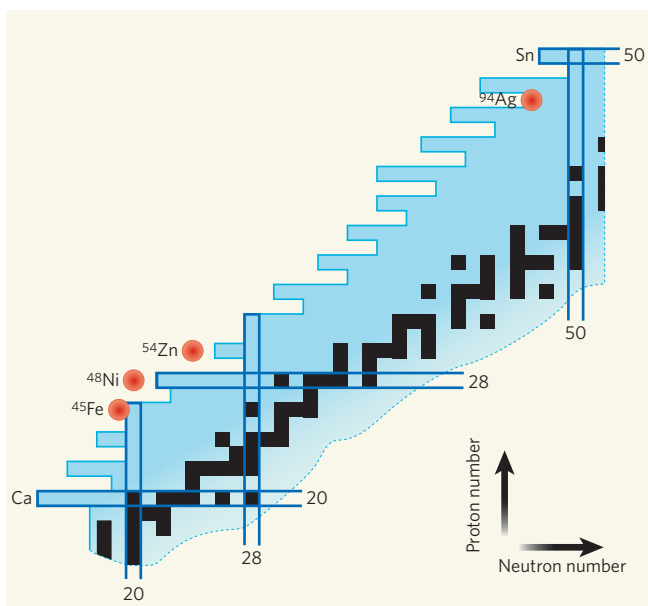


Figure 1 | Rich in proton. A part of the nuclide chart displaying proton-rich nuclei between calcium (Ca) and tin (Sn). The isotopes framed by the jagged solid line at the top (the 'proton drip-line') are predicted to be bound (that is, stable against proton emission). The black squares denote stable isotopes; proton and neutron numbers 20, 28 and 50 are 'magic' nucleon numbers for which nuclei are especially stable⁷. The isotopes with recently discovered two-proton radioactivities are shown by red circles. Three of them, ^{45}Fe , ^{48}Ni and ^{54}Zn (refs 4, 5), are outside the proton drip-line and decay through two-proton radioactivity from the unbound ground state. In the case of ^{94}Ag , the ground state is bound against one-proton and two-proton emission — but the high-spin state investigated by Mukha *et al.*¹ is unbound for these decay modes.

of the protons was in all cases simultaneous; because of the extra energy required to break proton pairs, sequential emission is not energetically possible. A quantum-mechanical tunnelling model involving three bodies — two protons and a remnant core nucleus — describes the observations satisfactorily⁶, but for the process to be further elaborated theoretically, more precise information is required about the lifetime and decay energy of the system, and about the energy and angular correlations between the emitted protons.

The two-proton radioactivity observed by Mukha and colleagues¹ was from ^{94}Ag nuclei, not in their ground state, but in a metastable, long-lived, high-spin state that the authors produced by bombarding a nickel-58 target with a beam of calcium-40 ions. They also contrived to measure the energy and angular correlations of a two-proton decay for the first time; the proton-proton and proton-nucleus correlations derived are characteristic of simultaneous

rather than sequential emission. As ^{94}Ag was already known to undergo single-proton decay, Mukha and colleagues' state is the first for which both one- and two-proton decay modes have been shown to exist.

All previous studies of the two-proton decay of short-lived nuclear resonances from which sequential decay is energetically possible have proved that these decays are indeed exclusively sequential. Sequential emission is probably suppressed in ^{94}Ag because the daughter state that is populated by the first of two sequential decays would be an excited state of the palladium isotope ^{93}Pd . Rather than emitting a second proton, this state would tend to decay electromagnetically — by emitting a photon — to its ground state. But why does the energetically unfavourable simultaneous two-proton decay itself occur with such high probability?

Mukha *et al.* tackle this question using a simple model¹ that neglects any interaction between protons, but that is consistent with the observed energy and angular correlations as well as the half-life of the ^{94}Ag nucleus. Their model describes the decay as two independent emissions at either end of a strongly prolate (cigar-like) super-deformed nucleus. The authors supply a similar calculation that assumes the emission of a correlated proton pair with a relative spin of zero (a so-called ^2He model). However, the small number of events, and consequent limited accuracy of the results, does not allow the two modes to be distinguished. Further work is also needed to independently characterize the structure and shape of the complex high-spin ^{94}Ag state and thus allow a more detailed interpretation of the process.

Two-proton radioactivity can provide far-reaching insights into problems of low-energy quantum-mechanical tunnelling in strongly interacting systems. A full elucidation of the nature of such decays — whether from the ground state or from fast-spinning excited states — will be a major goal for further experiments at current and future radioactive ion-beam facilities.

Juha Äystö is in the Department of Physics, University of Jyväskylä, Jyväskylä 40351, Finland. e-mail: juha.aysto@phys.jyu.fi

1. Mukha, I. *et al.* *Nature* **439**, 298–302 (2006).
2. Jackson, K. P. *et al.* *Phys. Lett. B* **33**, 281–283 (1970).
3. Woods, P. J. & Davids, C. N. *Annu. Rev. Nucl. Part. Sci.* **47**, 541 (1997).
4. Dossat, C. *et al.* *Phys. Rev. C* **72**, 054315 (2005).
5. Blank, B. *et al.* *Phys. Rev. Lett.* **94**, 232501 (2005).
6. Grigorenko, L. V. & Zhukov, M. V. *Phys. Rev. C* **68**, 054005 (2003).
7. Janssens, R. V. F. *Nature* **435**, 897–898 (2005).

OBITUARY

Theodore H. Bullock (1915–2005)

Trailblazer in neurobiology.

Last summer, Ted Bullock and I walked under the wooden arcade covered by bougainvillea vines at Casa de Mañana, the retirement community in La Jolla, southern California, where he lived with his wife. All of a sudden, Ted stopped and pointed to the rich pattern of light and shade on the ground from the sun penetrating the plants. "Have you noticed that the size of the light points enclosed by the shadows are never smaller than a certain minimum size?" he said, and expounded the physical reasons he thought might cause this "quantum" phenomenon.

For Ted Bullock, who died in the night of 19–20 December 2005, no problem was too trivial to be examined, no question too simple to be thought about. This never-ending curiosity was key to his success as one of the greatest neurobiologists of the twentieth century. Like no one before him, he pioneered a reductionist approach based on careful observations of a large variety of animal species under natural conditions to explore how nervous systems work, how they control behaviour, and how they developed during evolution.

Theodore Holmes Bullock was born to American Presbyterian missionaries in Nanking, China, in 1915. During his first 13 years, immersion in the Chinese culture, with its atmosphere of mutual respect, laid the foundation for his lifelong cosmopolitan outlook. He considered himself, "a citizen of the world first, of the United States second". After earning his PhD in zoology from the University of California at Berkeley in 1940, he spent four years at Yale University in New Haven, Connecticut, with summers at the Marine Biology Laboratory at Woods Hole, Massachusetts.

There, the ample opportunity to work on a large variety of organisms had a major influence on the development of Bullock's interests in comparative physiology. He continually searched for simple systems with which to explore the multitude of neural mechanisms that work together to produce an output, both at the physiological and the behavioural level. He argued that examination of the different mechanisms used by various organisms is as important as the search for commonalities in understanding how brains work. As a consequence, he studied an enormous number of different taxa over his lifetime, including cnidarians, crustaceans, fish, amphibians and reptiles.

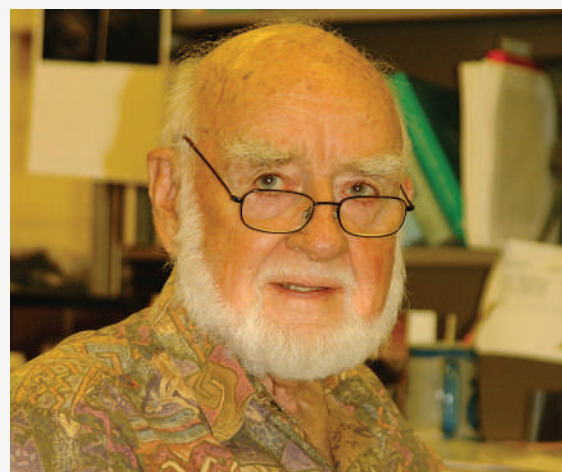
His research life entered a decisive stage when he was appointed to the zoology faculty

of the University of California at Los Angeles in 1946. Twenty years later, he moved to the University of California at San Diego, where he headed the Neurobiology Unit of the Scripps Institution of Oceanography and became a major force behind the creation of the world's first department of neuroscience.

One of the research themes to which Bullock returned time and again, was the role of direct current and low-frequency electric fields in intercellular communication in the nervous system. In the early 1960s, the dominant view of neuronal communication was that it takes place only through action potentials, with chemical synapses acting as switches to regulate the flow of information through neural circuits. Bullock, however, showed that neurons can also communicate through non-synaptic interactions, and without such impulses. Using the cardiac ganglion of lobsters as a particularly favourable model system, he demonstrated that graded potentials spread electrically from one neuron to another through proposed low-resistance cytoplasmic bridges. Today we know that this type of electrical interaction is mediated by gap junctions — electrical synapses that couple groups of cells into functional units.

As an extension of this work, Bullock maintained a lifelong interest in field potentials resulting from the summated electrical activity of millions of brain cells. He argued that present approaches to analyse these electric activities are insufficient to extract all of the information possibly used for signalling within the brain. Furthermore, he demonstrated that certain patterns of field potentials underlying cognitive processes can be recorded not only from higher brain centres in taxonomically higher vertebrates, but also from sites involved in low-level sensory processing (for example, the retina) and in lower vertebrates, such as rays and turtles. On the basis of Bullock's research, this comparative approach has proven to be a powerful tool for uncovering principles of the evolution of cognition.

Bullock's other scientific contributions include the discovery of two new senses in animals. In a series of papers in the 1950s, he and his associates showed that the facial pit of pit vipers, located midway between the nostril and eye on either side of the head, acts as a thermal sensor. The snakes use receptors in the facial pit to detect temperature changes caused by prey animals in the vicinity. A second landmark discovery



came in the early 1960s, when he and his colleagues found a novel class of sensory organ in the skin of certain fish. Using a specialized organ, these fish produce weakly electric discharges that aid orientation and communication with members of their own species. The receptor cells of the electric sense organs respond to electric fields, even of extremely low intensities, and are tuned optimally to detect the fish's own discharges and those of neighbouring electric fish.

A researcher who maintained the highest standards, Ted remained enormously productive until his death. He was an inspiring teacher who encouraged broad education and independence. Over his lifetime, he attracted a tremendous number of graduate students, postdocs and visiting investigators from all over the world, helping many of them to enter promising research areas, without necessarily pursuing the field himself. Above all, his attitude towards people was clearly that of a humanist. He was dedicated to his wife of 68 years, Martha, and their two children. He always had time for other people and, when asked, gave advice — not by dictating what to do, but by presenting a well-balanced analysis.

The volume of his correspondence with friends, students and colleagues all over the world was immense and hardly decreased with age — as the photo above from last year shows, he remained youthful and vigorous even at 90. His last e-mail to me was sent just a few hours before his death, with detailed suggestions on a manuscript. He, who considered it a duty and a pleasure to help others, finished his e-mail "thanks for the privilege". It was a privilege to be among those with whom he shared his wisdom, enthusiasm and love for science, nature and people.

Günther K. H. Zupanc

Günther K. H. Zupanc is at the School of Engineering and Science, International University Bremen, PO Box 750 561, D-28725 Bremen, Germany.
e-mail: g.zupanc@iu-bremen.de

G. K. H. ZUPANC

BRIEF COMMUNICATIONS

Superplastic carbon nanotubes

Conditions have been discovered that allow extensive deformation of rigid single-walled nanotubes.

The theoretical maximum tensile strain — that is, elongation — of a single-walled carbon nanotube is almost 20%^{1,2}, but in practice only 6%^{3,4} is achieved. Here we show that, at high temperatures, individual single-walled carbon nanotubes can undergo superplastic deformation, becoming nearly 280% longer and 15 times narrower before breaking. This superplastic deformation is the result of the nucleation and motion of kinks in the structure, and could prove useful in helping to strengthen and toughen ceramics and other nanocomposites at high temperatures.

A single-walled carbon nanotube (SWCNT) with an initial length of 24 nm (Fig. 1a) was formed *in situ* by the electrical breakdown of a multiwalled carbon nanotube inside a high-resolution transmission electron microscope equipped with a piezo manipulator⁵. The piezo manipulator was used to pull the SWCNT to increase the strain (Fig. 1b–d), at a constant bias of 2.3 volts. At tensile failure, the SWCNT was 91 nm long, showing a tensile elongation of 280%; its diameter was reduced 15-fold, from 12 to 0.8 nm. As the SWCNT segment was clamped between the layers of the multiwalled carbon nanotube, the possibility of wall-sliding during elongation is ruled out. The diameter of the elongated nanotubes did not change significantly when wall-sliding occurred.

A 280% tensile strain and 15-fold reduction in diameter are unprecedented in a SWCNT.

Ropes of single-walled^{3,4} and multiwalled⁶ carbon nanotubes break at strains of less than about 6% and 12%, respectively, which is significantly less than that seen here. We propose that the super-elongation we observe is due to a fully plastic deformation mechanism that occurs at high temperatures.

During deformation at a bias of 2.3 V, the temperature in the middle of the SWCNT is estimated to be more than 2,000 °C (see supplementary information). At such high temperatures, the nanotube seems to be completely ductile. Kinks and point defects are fully activated, resulting in superplastic deformation impossible at low temperatures. Kinks form frequently during tensile straining (Fig. 1b–d), propagate along the tube and then pile up (Fig. 1d) or disappear at the ends (for movie, see supplementary information).

The kink motion is evidence of kink-mediated plasticity at high temperatures. The nanotube narrows immediately after the kink passes. We suggest that these kinks are associated with one or several unit dislocations that have a Burgers vector of $1/3\langle 1120 \rangle$ type.

Such large plastic strains in nanotubes demonstrate their ductile nature at high temperatures^{7–9}. In contrast, tensile-pulling experiments at room temperature without any bias showed that almost all nanotubes failed at a tensile strain of less than 15% (Fig. 1e–g). Our experimental evidence (not shown) shows

that, as well as kink nucleation and motion, atom diffusion is important during superplastic deformation, helping to heal defects such as vacancies and to prevent the formation of large rings that might initiate cracks and lead to failure of the strained nanotube.

When the diameter and length of a SWCNT change in this dramatic way, the current flowing in the nanotube drops sharply from 80 μA to almost zero, causing the current density to drop from 10^9 to 10^8 A cm⁻² and then to almost zero (for details, see supplementary information). An increase in defect density causes this rapid fall in current density as the nanotube diameter decreases.

Our surprising discovery of superplasticity in nanotubes should encourage the investigation of their mechanical and electronic behaviour at high temperatures. They may be useful as reinforcement agents in ceramics and other nanocomposites for high-temperature applications^{9–11}.

J. Y. Huang*, S. Chen*, Z. Q. Wang*, K. Kempa*, Y. M. Wang†, S. H. Jo*, G. Chen‡, M. S. Dresselhaus§, Z. F. Ren*

*Department of Physics, Boston College, Boston, Massachusetts 02467, USA

e-mail: huangje@bc.edu

†Lawrence Livermore National Laboratory, Nanoscale Synthesis and Characterization Laboratory, Livermore, California 94550, USA

‡Departments of §Mechanical Engineering, and §Physics, Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

1. Zhao, Q. Z., Nardelli, M. B. & Bernholc, J. *Phys. Rev. B* **65**, 144105 (2002).
2. Nardelli, M. B., Yakobson, B. I. & Bernholc, J. *Phys. Rev. B* **57**, R4277–R4280 (1998).
3. Walters, D. A. *et al. Appl. Phys. Lett.* **74**, 3803–3805 (1999).
4. Yu, M. F., Files, B. S., Arepalli, S. & Ruoff, R. S. *Phys. Rev. Lett.* **84**, 5552–5555 (2000).
5. Huang, J. Y. *et al. Phys. Rev. Lett.* **94**, 236802 (2005).
6. Yu, M. F. *et al. Science* **287**, 637–640 (2000).
7. Nardelli, M. B., Yakobson, B. I. & Bernholc, J. *Phys. Rev. Lett.* **81**, 4656–4659 (1998).
8. Orlikowski, D., Nardelli, M. B., Bernholc, J. & Roland, C. *Phys. Rev. Lett.* **83**, 4132–4135 (1999).
9. Calvert, P. *Nature* **399**, 210–211 (1999).
10. Zhan, G. D., Kuntz, J. D., Wan, J. L. & Mukherjee, A. K. *Nature Mater.* **2**, 38–42 (2003).
11. Baughman, R. H., Zakhidov, A. A. & de Heer, W. A. *Science* **297**, 787–792 (2002).

Supplementary information accompanies this communication on Nature's website.

Competing financial interests: declared none.
doi:10.1038/439281a

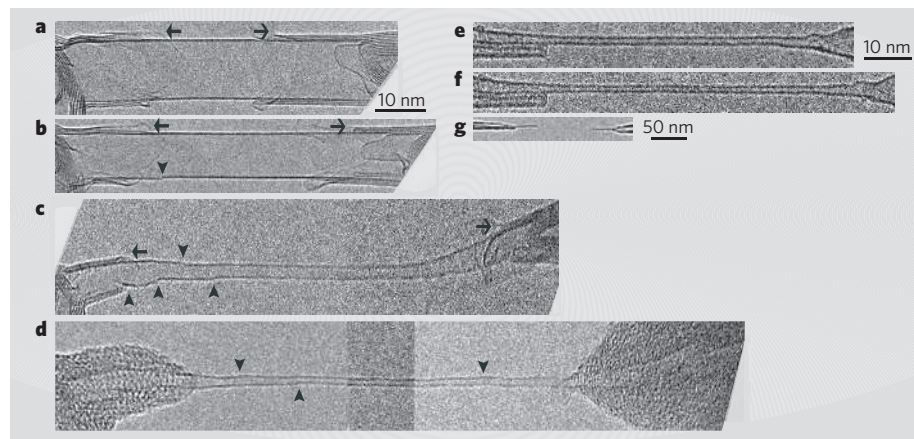


Figure 1 | *In situ* tensile elongation of individual single-walled carbon nanotubes viewed in a high-resolution transmission electron microscope. **a–d**, Tensile elongation of a single-walled carbon nanotube (SWCNT) under a constant bias of 2.3 V (images are all scaled to the same magnification). Arrowheads mark kinks; arrows indicate features at the ends of the nanotube that are almost unchanged during elongation. **e–g**, Tensile elongation of a SWCNT at room temperature without bias (images **e** and **f** are scaled to the same magnification). Initial length is 75 nm (**e**); length after elongation (**f**) and at the breaking point (**g**) is 84 nm; **g**, low-magnification image of the SWCNT breaking in the middle.

A brain-specific microRNA regulates dendritic spine development

Gerhard M. Schratt^{1,2,3}, Fabian Tuebing⁴, Elizabeth A. Nigh^{1,2,3}, Christina G. Kane^{1,2,3}, Mary E. Sabatini³, Michael Kiebler⁴ & Michael E. Greenberg^{1,2,3}

MicroRNAs are small, non-coding RNAs that control the translation of target messenger RNAs, thereby regulating critical aspects of plant and animal development. In the mammalian nervous system, the spatiotemporal control of mRNA translation has an important role in synaptic development and plasticity. Although a number of microRNAs have been isolated from the mammalian brain, neither the specific microRNAs that regulate synapse function nor their target mRNAs have been identified. Here we show that a brain-specific microRNA, miR-134, is localized to the synaptodendritic compartment of rat hippocampal neurons and negatively regulates the size of dendritic spines—postsynaptic sites of excitatory synaptic transmission. This effect is mediated by miR-134 inhibition of the translation of an mRNA encoding a protein kinase, *Limk1*, that controls spine development. Exposure of neurons to extracellular stimuli such as brain-derived neurotrophic factor relieves miR-134 inhibition of *Limk1* translation and in this way may contribute to synaptic development, maturation and/or plasticity.

Highly orchestrated programmes of gene expression act to shape the developing nervous system. This tight regulation is mediated by a variety of transcriptional and post-transcriptional events that control the expression of individual gene products^{1,2}. The discovery of small, non-coding RNAs has greatly expanded our understanding of the cellular mechanisms that regulate gene expression at the post-transcriptional level. MicroRNAs (miRNAs) act by binding to target mRNAs and initiating either cleavage or a reduction in the translational efficiency of the target mRNA, depending on the degree of sequence complementarity^{3–5}. Biochemical and genetic studies have revealed important functions for specific miRNAs in a variety of cellular processes, including differentiation, apoptosis and metabolism^{6–10}.

A number of miRNAs have been isolated from the vertebrate nervous system^{11–13}, and a recent study has demonstrated a crucial role for the miRNA pathway in early zebrafish brain development¹⁴. Expression analysis also supports a role for miRNAs in later stages of neuronal maturation and synapse development^{12,15,16}. A potential role for miRNAs in synaptic function is particularly intriguing given the evidence that selected mRNAs in neurons are transported to sites of synaptic contact that are quite distant from the cell body^{17–19}. Within dendrites, and at synapses, the translation of these mRNAs may be inhibited until neurons are exposed to appropriate extracellular stimuli such as a neurotrophic factor (for example, brain-derived neurotrophic factor (BDNF)) or neurotransmitter release at the synapse. Local translation of these previously dormant mRNAs has been hypothesized to have a key role in synaptic development and plasticity^{20–22}. Whether miRNAs might inhibit the translation of synaptically localized mRNAs in neurons until their translation is activated by neurotrophic factors or neuronal activity remains to be investigated.

miR-134 expression during synapse development

To identify miRNAs that might function in dendritic and/or synaptic development, we investigated the expression and localization of

candidate miRNAs that had been previously isolated from mouse brain¹³. Both northern blotting and an RNase protection assay (RPA) revealed that the expression of microRNA-134 (miR-134) is restricted to the brain, similar to the expression pattern of the previously characterized miR-124a (Fig. 1a and Supplementary Fig. 1a, b). Unlike miR-124a, however, miR-134 levels in the hippocampus gradually increase with development, reaching maximum levels at postnatal day 13 (P13), the time at which synaptic maturation occurs (Fig. 1b). A similar developmental expression profile was also observed in dissociated hippocampal neurons that were allowed to mature over time in culture (Fig. 1c). Moreover, membrane depolarization of cortical neurons induced a significant increase in the level of the miR-134 precursor (Supplementary Fig. 1c). Taken together, these results suggested a potential role for miR-134 in dendritic and/or synaptic development.

We used an *in situ* hybridization (ISH) protocol to examine the subcellular localization of the miR-134 RNA within cultured hippocampal neurons. Unlike the mismatch control probe, hybridization with the miR-134-specific probe revealed the presence of miR-134 within dendrites, where it is present in a punctate pattern (Fig. 1d and Supplementary Fig. 1d). Quantification of the two signal intensities (miR-134-specific versus the mismatch probe) along the length of multiple dendrites confirmed significantly higher levels of miR-134-specific signal within dendrites as compared to that obtained with the mismatch control (Supplementary Fig. 1e) or the U6 small nuclear (sn)RNA (data not shown). A substantial fraction of the dendritic miR-134 was found to partially co-localize with synapsin immunostaining, indicating that miR-134 is present near synaptic sites on dendrites (Fig. 1d, lower panel and inset at higher magnification). The presence of miR-134 in synaptic compartments was also corroborated by subcellular fractionation experiments; miR-134 was enriched in synaptoneurosome preparations (Fig. 1e), which represent membrane preparations highly enriched for synaptic terminals²³. The presence of miR-134 within dendrites near synapses suggested a possible functional role for this miRNA at post-synaptic sites.

¹Neurobiology Program, Children's Hospital, ²Department of Neurology, ³Department of Neurobiology, Harvard Medical School, Boston, Massachusetts 02115, USA. ⁴Division of Neuronal Cell Biology, Center for Brain Research, Medical University of Vienna, A-1090 Vienna, Austria.

miR-134 regulates dendritic spine morphology

To investigate a possible function of miR-134 at the synapse, we examined the effects of modulating miR-134 activity on dendritic spine development. Dendritic spines are actin-rich protrusions from the dendritic shaft and represent the major sites of excitatory synaptic contact^{24,25}. The size of dendritic spines is a good correlate of the strength of excitatory synapses^{26–28}. To achieve miR-134 overexpression, we designed a vector that permits efficient expression of exogenous miR-134 (Supplementary Fig. 2a). Alternatively, miR-134 function in neurons was suppressed by introducing a 2'-O-methylated antisense oligonucleotide that interferes with endogenous miR-134 activity in a sequence-specific manner^{29,30}. The efficacy of these approaches was confirmed in neurons using a previously described miRNA sensor assay (Supplementary Fig. 2b)³¹.

An analysis of dendritic spines in cultured hippocampal neurons (cultured for a total of 18 days, transfected at day 8: 8 + 10 days *in vitro* (DIV)) overexpressing miR-134 showed a significantly decreased spine volume as compared to the spines of neurons transfected with empty vector or overexpressing the unrelated let-7c miRNA (Fig. 2a (bottom panel), b (bottom panel), e and Supplementary Fig. 3a–c).

Further analysis revealed that this decrease in spine volume was mainly a consequence of a reduction in spine width ($-16.9 \pm 5.8\%$, $n = 3$, $P = 0.02$) as opposed to a change in spine length ($-3.5 \pm 7.1\%$, $n = 3$, $P = 0.23$, Fig. 2f). A similar reduction in

dendritic spine size was observed when synthetic miR-134 was introduced into neurons at a later stage (15 DIV) and for shorter times (72 h, Supplementary Fig. 3d). Because hippocampal neurons at 15 DIV have already developed the vast majority of their spines, these findings suggest that miR-134 may perturb the morphology of pre-existing spines.

In contrast to the effect of miR-134 overexpression, sequence-specific inhibition of endogenous miR-134 function using a 2'-O-methylated antisense oligonucleotide^{29,30} (2'-O-Me-134) led to small but statistically significant increases in spine volume and width ($7.6 \pm 3.7\%$, $n = 3$, $P = 0.03$) when compared to neurons transfected with an unrelated 2'-O-Me-control oligonucleotide (Fig. 2c (bottom panel), d (bottom panel), e, f). No significant effects on spine length were observed in the presence of 2'-O-Me-134 ($2.6 \pm 5.6\%$, $n = 3$, $P = 0.25$). Neither miR-134 overexpression nor the use of 2'-O-methylated oligonucleotides had any measurable effect on spine density or overall dendritic complexity (Supplementary Fig. 3e, f). We conclude that miR-134 acts as a negative regulator of dendritic spine volume in hippocampal neurons, raising the possibility that miR-134 may be involved in the regulation of synapse development and/or function.

miR-134 inhibits translation of *Limk1* mRNA

To gain insight into the mechanisms by which miR-134 regulates dendritic spine morphology, we sought to identify miR-134 target mRNAs. Towards this end, we scanned the 3' untranslated regions (UTRs) of mRNAs for potential miR-134 binding sites. For this analysis, we focused on a set of 48 genes that we recently identified in a screen for mRNAs for which translation is enhanced in neurons upon treatment with BDNF³². As BDNF promotes dendritic spine growth³³ and regulates synaptic function, at least in part, by activating dendritic protein synthesis²¹, we reasoned that mRNAs for which translation is regulated by BDNF might also represent miR-134 targets. Three of the BDNF-regulated mRNAs (discs large homologue 2 (*DLG2*), *Neurod2* and Lim-domain-containing protein kinase 1 (*Limk1*)) were found to contain conserved 3' UTR sequence elements that were partially complementary to mouse miR-134 (Fig. 3a and data not shown). Among these potential miR-134 target mRNAs, *Limk1* was of particular interest. *Limk1* regulates actin filament dynamics through inhibition of ADF/cofilin³⁴, and *Limk1* knockout mice show abnormalities in dendritic spine structure similar to those observed upon miR-134 overexpression³⁵.

Using an electrophoretic mobility shift assay, we demonstrated that *Limk1* mRNA and miR-134 interact *in vitro* (Supplementary Fig. 4). We next determined whether the *Limk1* mRNA co-localized with miR-134 within dendrites of live neurons. Fluorescently labelled miR-134 was introduced into hippocampal neurons by micro-injection together with a fluorescent *Limk1* 3' UTR that contains the miR-134 binding site. Both miR-134 and the *Limk1* 3' UTR, in contrast to the non-dendritic *Gapdh* and histone H3 mRNAs, were found to be present within dendrites in a granular pattern (Fig. 3b, upper two panels and data not shown). Moreover, miR-134- and *Limk1* 3' UTR-positive granules were co-localized within dendrites (Fig. 3b, lower panel). Furthermore, efficient co-localization of miR-134 and *Limk1* mRNA required the presence of an intact miR-134 binding site within the *Limk1* 3' UTR (Fig. 3b, bar graph). The dendritic localization of endogenous *Limk1* mRNA was further confirmed by ISH in cultured neurons and by subcellular fractionation (Supplementary Fig. 5a, b).

In mammalian cells, miRNAs are thought to regulate the expression of target mRNAs predominantly through the inhibition of productive translation³. We therefore hypothesized that miR-134 binding to the *Limk1* mRNA might act to inhibit *Limk1* translation. In support of this idea, miR-134 overexpression in both 293T cells and primary neurons was found to decrease specifically the activity of a luciferase reporter gene fused to the wild-type *Limk1* 3' UTR, whereas expression of the unrelated let-7c miRNA had no significant

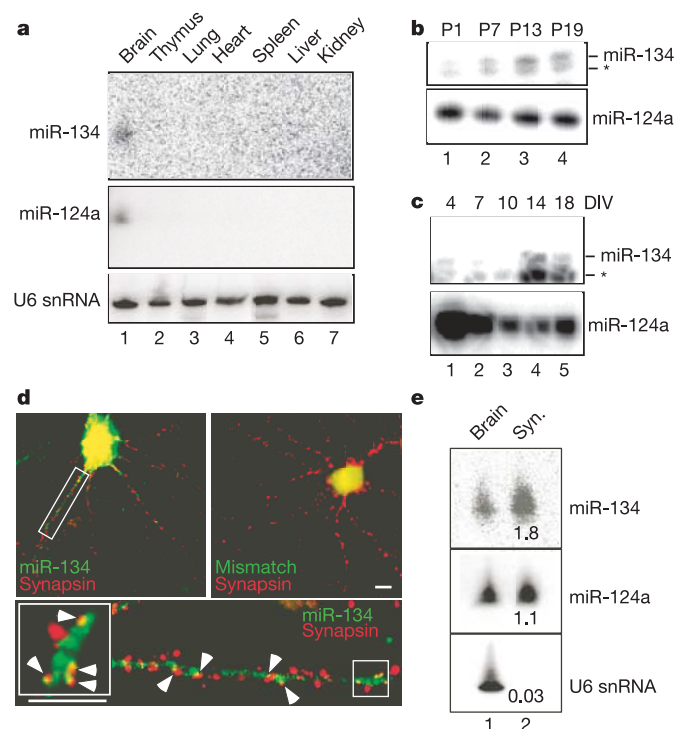


Figure 1 | miR-134 is specifically expressed in the brain and localized to neuronal dendrites. **a**, Northern blot of adult tissues was probed for the indicated miRNAs or U6 snRNA. **b**, RNase protection assay (RPA) to detect the indicated miRNAs in postnatal (P1–P19) hippocampus. **c**, RPA to detect the indicated miRNAs in hippocampal neurons cultured for 4–18 DIV. Asterisk indicates an unknown protected fragment. **d**, Co-staining of the presynaptic marker protein synapsin (red) together with miR-134 ISH (green) in 14 DIV hippocampal neurons (upper left). The boxed area in the upper-left panel is shown at greater magnification in the bottom panel, which also has a higher-magnification inset. Arrows point to synapses that partially overlap with miR-134-positive puncta. Scale bars, 10 μ m. **e**, Northern blot of P15 whole brain or synaptoneurosomes (syn.) probed for indicated miRNAs or U6 snRNA. Fold enrichment in synaptoneurosomes is depicted.

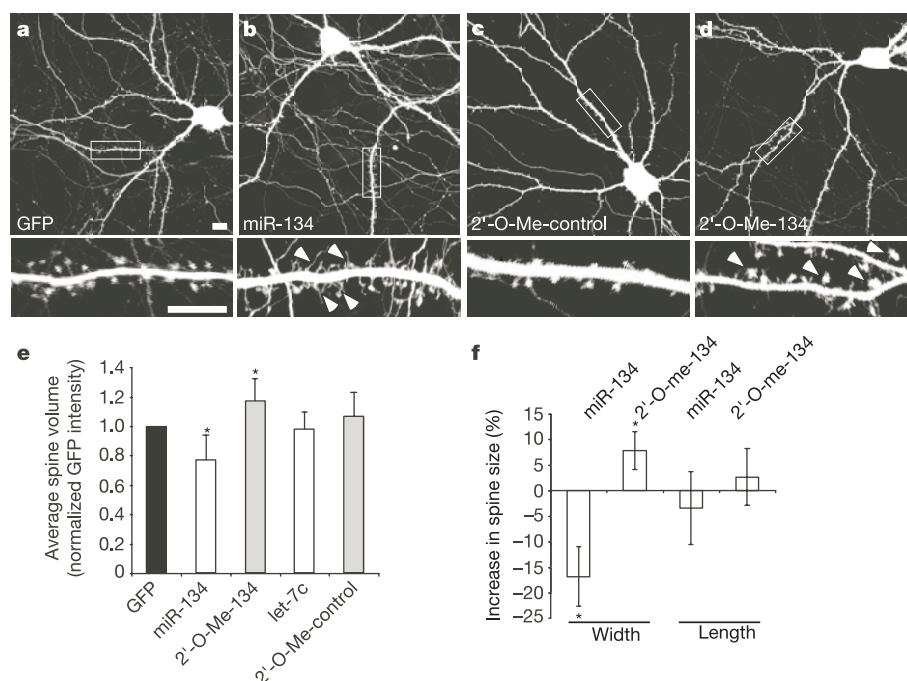


Figure 2 | miR-134 negatively regulates dendritic spine size in hippocampal neurons. **a–d**, Representative neurons (18 DIV) transfected with control vector (**a**), miR-134 expression vector (**b**), 2'-O-Me control (**c**) or 2'-O-Me-134 oligonucleotide (**d**). Bottom panels (insets of boxed areas) illustrate higher frequency of thinner spines in miR-134-expressing cells (arrows in **b**) and enlarged spines in 2'-O-Me-134-transfected neurons (arrows in **d**). Scale bars, 10 μ m. **e**, Normalized average volume of spines ($n > 600$) from

neurons ($n = 15$) transfected as in **a–d**. Data are presented as average spine volume $\pm$ s.d. from three independent experiments. Asterisk, $P < 0.05$ (paired Student's t -test). **f**, Per cent changes in the width and length of spines ($n > 600$) in neurons ($n = 15$) expressing miR-134 or 2'-O-Me-134 compared to GFP. Data are presented as mean change in spine length/width $\pm$ s.d. from three independent experiments. Asterisk, $P < 0.05$ (paired Student's t -test).

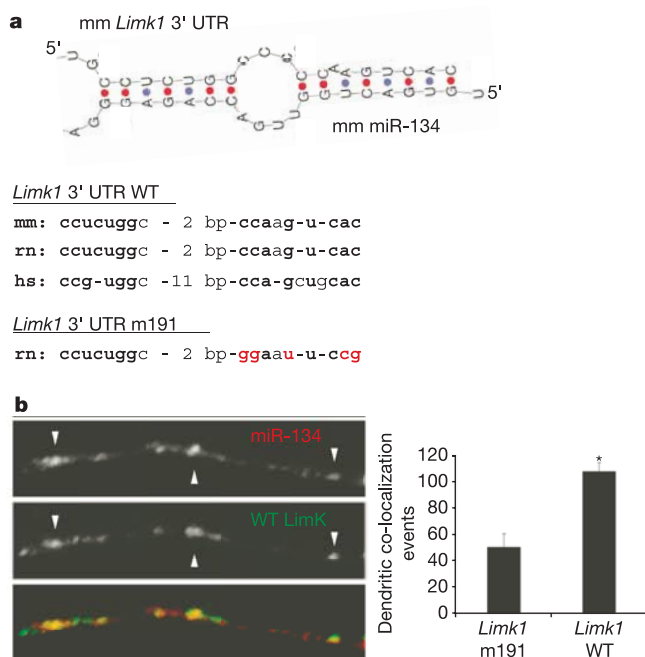


Figure 3 | *Limk1* mRNA is a putative miR-134 target. **a**, Upper panel: predicted duplex formation between mouse *Limk1* 3' UTR (top) and miR-134 (bottom). Middle panel: Sequence conservation of the miR-134 binding site within the *Limk1* 3' UTR of mouse (mm), rat (rn) and human (hs). Lower panel: sequence of the m191 *Limk1* 3' UTR, containing mutations in the miR-134 binding site (red). **b**, Localization of microinjected miR-134 (red) and *Limk1* RNA (green) in hippocampal neurons. Arrows indicate miR-134 and *Limk1* co-localization in granule-like structures. Bar graph: quantification of dendritic co-localization events between microinjected miR-134 and either *Limk1* m191 or *Limk1* wild-type (WT) RNA. Data represent the mean of $n = 12$ cells per condition counted in triplicate $\pm$ s.d. Asterisk, $P < 0.05$.

effect on the expression of this reporter construct (Fig. 4a and Supplementary Fig. 6a, b). The steady-state levels of the reporter gene mRNA were unaffected by miR-134 overexpression, suggesting that the observed effect of miR-134 on luciferase expression does not reflect a change in the stability of the luciferase mRNA (Supplementary Fig. 6c). The effect of miR-134 on translation of the luciferase mRNA is dependent on the presence of the miR-134 cognate binding site within the 3' UTR, as expression of a luciferase reporter containing the mutant m191 *Limk1* 3' UTR (that is, with a mutated miR-134 binding site) was unaffected by the presence of exogenous miR-134 (Fig. 4a, white bars). In contrast to the effect on *Limk1* mRNA translation, mutation of the miR-134 binding site did not affect dendritic targeting of a *Gfp-Limk1* reporter RNA (Supplementary Fig. 5c).

The inhibition of endogenous miR-134 in neurons by 2'-O-Me-134 led to a statistically significant increase in the expression of the luciferase reporter fused to the wild-type *Limk1* 3' UTR (Fig. 4b, black bars), but had no significant effect on expression of the m191 mutant reporter that is incapable of binding miR-134 (Fig. 4b, white bars). By contrast, an antisense oligonucleotide directed against let-7c (2'-O-Me-let-7c) had no effect on *Limk1* reporter gene activity. Peptide-mediated delivery of miR-134 into neurons led to a dose-dependent decrease in the level of endogenous *Limk1* protein, whereas delivery of its inhibitor 2'-O-Me-134 led to an increase in protein level (Fig. 4c, d), suggesting that miR-134 inhibits translation of the endogenous *Limk1* mRNA. Taken together, these data suggest that endogenous miR-134 inhibits *Limk1* mRNA translation in neurons by binding to a single site present in the *Limk1* 3' UTR.

Although these studies provide evidence that miR-134 acts to repress *Limk1* mRNA translation, they do not distinguish whether the inhibition occurs within the cell body and/or dendrites. To address this issue, we generated a GFP-based protein synthesis reporter (*myr-dlGfp*) with limited diffusion and a shortened half-life (1 h). Results from a previous study using a similar construct

demonstrated that GFP expressed from the reporter gene allows for the study of local protein synthesis within intact dendrites³⁶. The *myr-d1GFP* reporter was fused to either wild-type or m191 mutant *Limk1* 3' UTR and introduced into hippocampal neurons. GFP expression was monitored by confocal microscopy, and the intensity of the GFP signal was determined in the dendrites of many neurons at varying distances from the cell body (Fig. 4e). This analysis revealed that the average expression of the wild-type *Limk1* reporter was significantly reduced (by 18–28%) along the entire length of the dendrites compared to that of the m191 reporter (Fig. 4f). Given the dendritic localization of endogenous *Limk1* mRNA and miR-134, these findings suggest that miR-134 partially inhibits *Limk1* mRNA translation locally within dendrites.

miR-134 regulates spine size through *Limk1*

Because both overexpression of miR-134 and disruption of *Limk1* function lead to decreased spine size³⁵, we next investigated whether miR-134-mediated repression of *Limk1* mRNA translation might be an explanation for the observed reduction in dendritic spine size upon miR-134 overexpression. Towards this end, we expressed miR-134 in hippocampal neurons together with constructs expressing either a wild-type *Limk1* mRNA or mutant m191 *Limk1* mRNA, and monitored dendritic spine size (Fig. 5a). We reasoned that if the effect of miR-134 on spine morphology occurs through suppression of endogenous *Limk1* mRNA translation, ectopically expressed *Limk1* mRNA that is incapable of interacting with miR-134 (m191) should be able to rescue the spine defect. In contrast, co-expression of the wild-type *Limk1* mRNA, which is still subject to miR-134-mediated translational inhibition, might be expected to prove less effective in the rescue of the dendritic spine phenotype caused by miR-134

overexpression. Consistent with this idea, we found that the m191 mutant *Limk1* mRNA efficiently rescued both the spine volume and width decrease imposed by miR-134 overexpression, whereas the wild-type *Limk1* mRNA was not as effective at rescuing the decrease in spine volume and width (Fig. 5a, upper and lower left panels). Both *Limk1* constructs had no effect on dendritic spine length (Fig. 5a, lower right). In addition, the observed difference between the effect of the wild-type and m191 mutant *Limk1* mRNA on spine width was not due to intrinsic differences in the ability of the two mRNAs to be translated, because in the absence of miR-134, *Limk1* protein levels were equivalent in 293T cells transfected with the wild-type and mutant *Limk1* constructs (Fig. 5b). Immunohistochemistry revealed that in neurons, overexpressed *Limk1* protein was targeted to synaptic sites within spines (Fig. 5c), consistent with the possibility that an increased level of *Limk1* protein within spines might be responsible for the rescue of the spine morphology phenotype. Taken together, these results suggest that *Limk1* is a downstream effector of miR-134 in the control of dendritic spine development.

miR-134 functions in BDNF-stimulated *Limk1* synthesis

Dendritic mRNAs are transported to the synapto-dendritic compartment within RNA granules. During their transport and once they have arrived at synaptic sites, the translation of dendritic mRNAs may be suppressed until extracellular factors such as those released upon synaptic stimulation activate the translation of these dormant mRNAs^{17,19}. We asked whether the suppression of *Limk1* translation by miR-134 is relieved by extracellular stimuli such as BDNF. We first assessed whether the translation of *Limk1* mRNA is regulated by BDNF. Towards this end, synaptoneurosomes prepared from P15 rat brain were incubated with ³⁵S-methionine to label newly synthesized

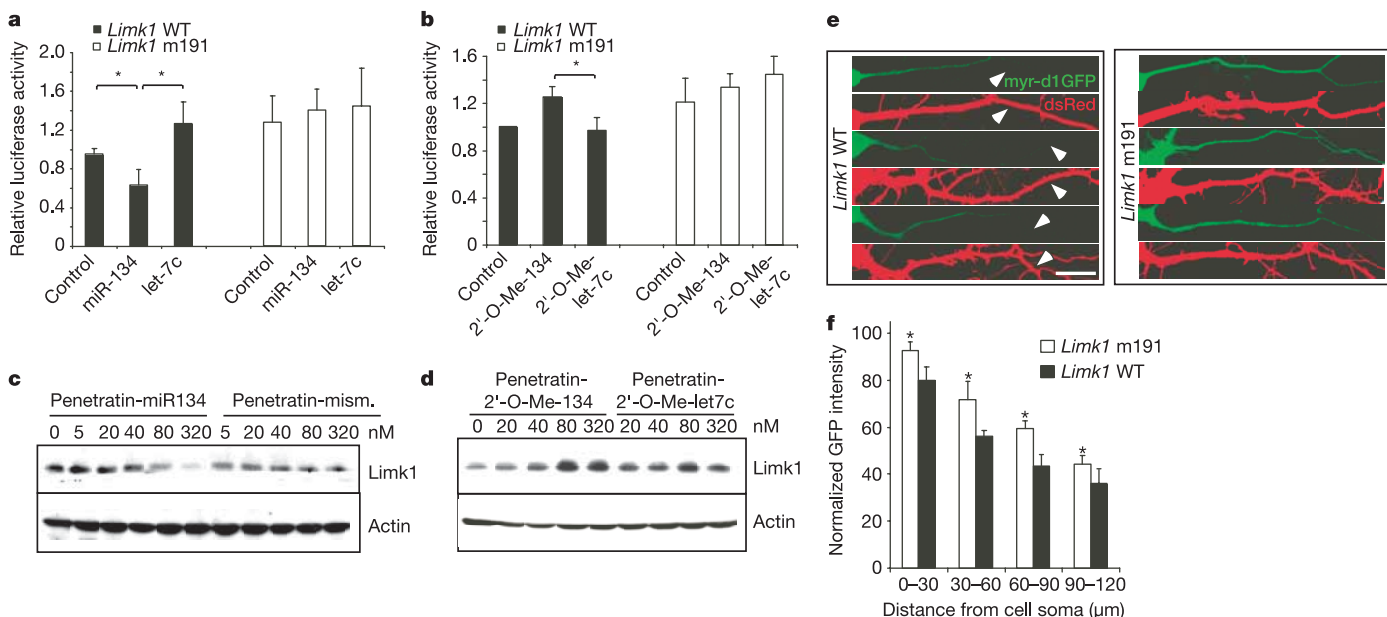


Figure 4 | miR-134 inhibits *Limk1* mRNA translation in neurons.

a, Luciferase activity of wild type (black bars) or m191 (white bars) *Limk1* 3' UTR reporter genes in the absence (control) or presence of the indicated miRNAs (10 μ M). Data represent the mean from three independent experiments $\pm$ s.d. Asterisk, $P < 0.05$ (paired Student's *t*-test). **b**, Luciferase activity of reporter genes described in **a** in the presence of the indicated antisense 2'-O-Me oligonucleotides (20 μ M). Data represent the mean from three independent experiments $\pm$ s.d. *Limk1* wild-type control = 1; asterisk, $P < 0.05$ (paired Student's *t*-test). **c**, Western blot analysis of endogenous *Limk1* (upper panel) and actin (lower panel) expression in lysates from cortical neurons (12 + 2 DIV) transduced with penetratin-coupled miR-134 or mismatch (mism.) control. **d**, Western blot analysis as in **c**, except that penetratin-coupled antisense 2'-O-Me oligonucleotides were used. **e**, Local

translation assay in hippocampal neurons (12 + 2 DIV) using destabilized, membrane-anchored *myr-d1GFP* reporter genes (green) harbouring either the wild type (left panel) or m191 (right panel) *Limk1* 3' UTR. Co-transfected dsRed was used to track dendrites. Three representative dendrites are shown per experimental condition. Arrows point to dendritic regions of *myr-d1GFP-Limk1* wild-type UTR transfected neurons where little GFP signal is detectable. Scale bar, 20 μ m. **f**, Average normalized GFP intensity in dendritic segments ($n > 60$) depicted in **e**. Data are from three independent experiments and presented as mean $\pm$ s.d. at 30- μ m dendritic intervals. The average GFP intensity of the *Limk1* m191 reporter at the most proximal part of the dendrite was set to 100. Asterisk, $P < 0.05$ (paired Student's *t*-test).

proteins, and the amount of newly synthesized Limk1 protein was monitored by radio-immunoprecipitation. BDNF treatment significantly increased synthesis of Limk1 protein within isolated synaptoneurosomes as indicated by an increase in ^{35}S -methionine-labelled protein in Limk1 immunoprecipitates. This increase was sensitive to treatment with rapamycin, an inhibitor of the mTOR kinase pathway, which we and others have shown to mediate BDNF signalling to the translational machinery (Fig. 6a)^{32,37}.

We next asked whether the ability of BDNF to induce *Limk1* mRNA translation reflects the ability of BDNF to relieve miR-134-dependent repression of *Limk1* translation. Towards this end, we examined the effect of BDNF treatment on the translation of a *Limk1* 3' UTR luciferase reporter mRNA in neurons at a time when endogenous miR-134 is highly expressed (14 DIV). When cells were transfected with luciferase mRNA fused to the wild-type *Limk1* 3' UTR, BDNF led to a statistically significant induction of

translation of the reporter mRNA (Fig. 6b). Expression of the m191 luciferase reporter mRNA was derepressed relative to the wild-type reporter in the absence of BDNF treatment, presumably due to the failure of endogenous miR-134 to bind to the m191 reporter gene (Fig. 6b). BDNF treatment did not lead to a further increase in the expression of the m191 reporter gene. To investigate the effect of miR-134 on BDNF-induced *Limk1* translation more directly, we introduced synthetic miR-134 into neurons that express little endogenous miR-134 (4 DIV, Fig. 1c). We found that miR-134 partially interferes with BDNF induction of the wild-type, but not the m191 mutant, reporter mRNA (Fig. 6c). These findings suggest that miR-134 represses *Limk1* mRNA translation and that BDNF treatment relieves this repression. However, the observation that there is still residual BDNF induction of reporter mRNA translation

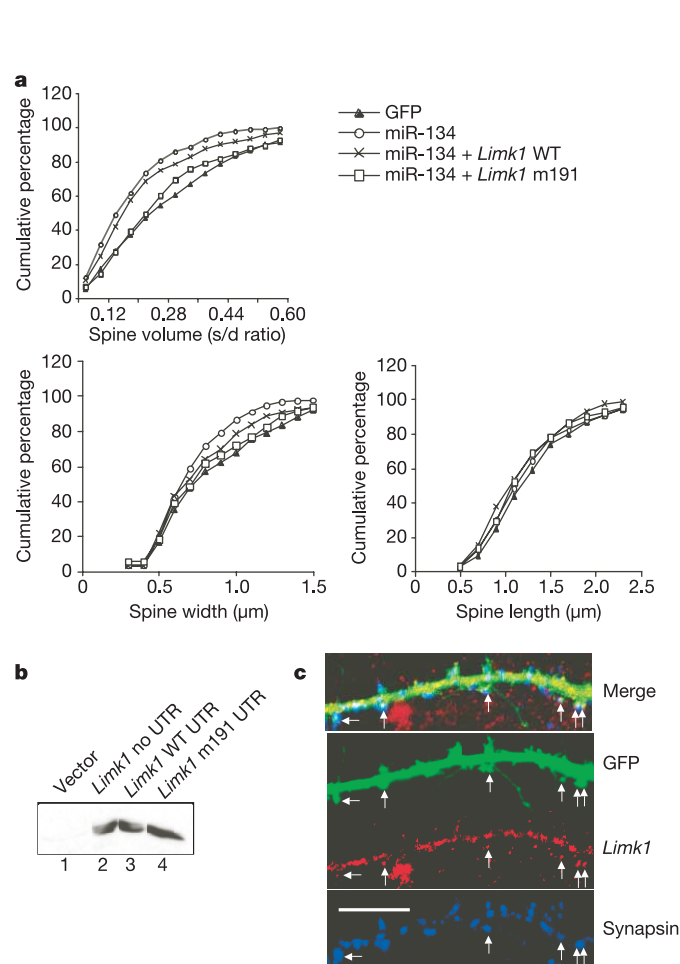


Figure 5 | *Limk1* expression rescues miR-134-mediated reduction in spine size. **a**, Cumulative percentage plots of spine volume, width and length in hippocampal neurons (18 DIV) transfected with miR-134 alone or together with the indicated *Limk1* expression constructs ($n > 500$ spines per condition from two independent experiments, five neurons per experiment). Spines of miR-134-transfected neurons have a significantly decreased volume compared to GFP ($P < 0.001$) or *Limk1* m191 3' UTR ($P < 0.001$)—but not *Limk1* wild-type 3' UTR ($P = 0.229$)—transfected neurons, and are significantly thinner than those of GFP ($P < 0.001$) or *Limk1* m191 3' UTR ($P = 0.006$)—but not *Limk1* wild-type 3' UTR ($P = 0.189$)—transfected neurons. Statistical significance was assessed by Kolmogorov–Smirnov test. **b**, Anti-Limk1 western blot of 293T whole-cell lysates transfected with vector alone or the indicated *Limk1* expression constructs. **c**, Immunocytochemistry of GFP (green), *Limk1* expressed from the *Limk1* m191 3' UTR construct (red) and synapsin (blue) in 18 DIV hippocampal neurons. Arrows point to the co-localization of *Limk1* and synapsin in GFP-positive dendritic spine heads. Scale bar, 10 μm .

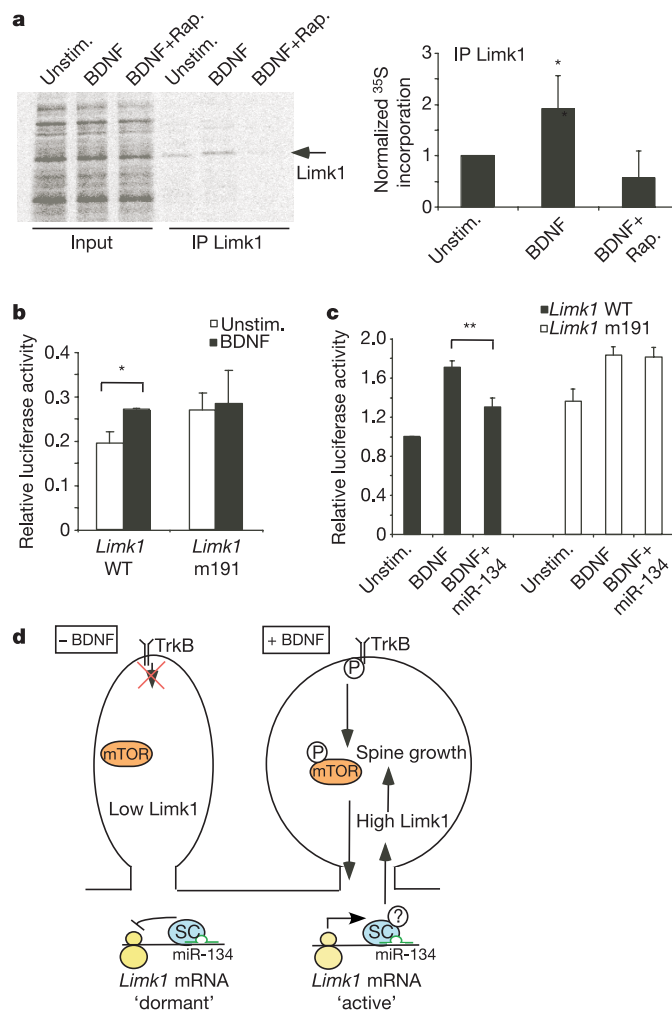


Figure 6 | miR-134 is involved in BDNF-induced *Limk1* mRNA translation. **a**, Left panel: immunoprecipitation (IP) of Limk1 from P15 synaptoneurosomes incubated with ^{35}S in the presence or absence of BDNF/rapamycin (rap). Right panel: average of the Limk1 immunoprecipitation signal intensities from three independent experiments $\pm$ s.d. Unstim. = 1. Asterisk, $P < 0.05$. **b**, Relative luciferase activity in 14 DIV cortical neurons transfected with *Limk1* wild type (black bars) or *Limk1* m191 (white bars) reporter mRNAs. Neurons were either unstimulated or treated with 100 ng ml $^{-1}$ BDNF for 4 h. Data represent the average of three independent experiments $\pm$ s.d. Asterisk, $P < 0.05$. **c**, Relative luciferase activity in 4 DIV cortical neurons transfected with *Limk1* wild type (black bars) or *Limk1* m191 (white bars) reporter mRNAs treated as in **b** together with miR-134 where indicated. Data represent the average of three independent experiments $\pm$ s.d. Asterisk, $P < 0.005$. **d**, Model for the role of miR-134 (green) in the regulation of *Limk1* synthesis and spine growth. For details, see text.

when miR-134 cannot bind to the *Limk1* 3' UTR suggests an involvement of additional miR-134-independent mechanism(s) in BDNF-induced *Limk1* translation.

Discussion

We have identified a dendritically localized miRNA that regulates the expression of the synaptic *Limk1* protein, thereby controlling dendritic spine size. We hypothesize that the association of *Limk1* mRNA with miR-134 keeps the *Limk1* mRNA in a dormant state while it is being transported within dendrites to synaptic sites (Fig. 6d). In the absence of synaptic activity, miR-134 may recruit a silencing complex that has a key role in repressing *Limk1* mRNA translation. This then limits the synthesis of new *Limk1* protein and restricts the growth of dendritic spines. Upon synaptic stimulation, the release of BDNF may trigger activation of the TrkB/mTOR signalling pathway, which inactivates the miR-134-associated silencing complex by an as-yet-unknown mechanism, leading to enhanced *Limk1* protein synthesis and spine growth. Our preliminary finding that miR-134 moves to the polysome-associated mRNA pool upon BDNF stimulation (G.S. and M.E.G., unpublished observations) suggests that miR-134 itself may not dissociate from the *Limk1* mRNA upon exposure of neurons to BDNF. Instead, we speculate that BDNF alters the activity of other translational regulators within the miR-134-containing complex. In addition to miR-134, other neuronal miRNAs have been predicted to bind the *Limk1* 3' UTR³⁸. Therefore, the combinatorial action of multiple miRNAs on the *Limk1* 3' UTR might explain our observation that miR-134 only partially inhibits *Limk1* mRNA translation (Fig. 4).

A recent bioinformatics approach predicted several additional neuronal mRNAs that may also represent miR-134 targets³⁹. Given that BDNF has important roles at multiple steps of synaptic development^{40,41}, it is possible that miR-134 regulates distinct sets of target genes involved in the formation, maturation or plasticity of synapses.

We propose that miRNA regulation of the translation of a variety of neuronal mRNAs will be found to contribute in an important way to synaptic function⁴². It is tempting to speculate that miRNAs act locally at individual synapses, thereby contributing to synapse-specific modifications that occur during synaptic plasticity. A future challenge will be to identify the full complement of dendritic miRNAs as well as their target mRNAs, and to determine their role in synaptic development.

METHODS

DNA constructs. The rat *Limk1* 3' UTR (1,171 base pairs) was amplified by polymerase chain reaction (PCR) from rat brain cDNA (P15). Mutation of the miR-134 binding site (m191) was achieved using the Quick Change site directed mutagenesis kit (Stratagene). PCR products were cloned into pGL3 basic (Promega), pBSK (Stratagene) or *myr-d1GFP* (gift of B. Sabatini) for constructs used in luciferase assay, *in vitro* transcription, or local reporter assay, respectively. For *Limk1* expression constructs, the *Limk1* cDNA (gift of K. Mizuno) was cloned into pcDNA3 (Promega) together with rat *Limk1* 3' UTR (wild type or m191). For the miR-134 expression construct, a genomic sequence spanning 150 base pairs 3' and 5' of the miR-134 sequence (Supplementary Fig. S1c) was PCR-amplified and cloned into pcDNA3. See Supplementary Information for further details.

Cell culture, transfection and stimulation. Cultures of dissociated primary cortical and hippocampal neurons were prepared as described³². Hippocampal neurons were maintained in Neurobasal plus B27 supplement; cortical neurons in Basal Medium Eagle plus 5% FBS. Neuronal transfections were performed with LipofectAmine 2000 (Invitrogen). For BDNF stimulation, neurons were starved overnight in the presence of UO126 (1 μ M) and then treated with BDNF (Preprotech, 100 ng ml⁻¹) for 4 h before cell harvest.

Northern blotting and RNase protection assays. RNA was isolated from synaptoneurosomes or cultured neurons by phenol/chloroform extraction using RNA Stat-60 (Tel-Test). For northern blots, 30 μ g of total RNA was resolved on 15% urea/polyacrylamide gels and transferred to Hybond N⁺ membrane (Amersham). See Supplementary Information for further details. RNase protection assays were performed with the mirVana miRNA detection kit (Ambion) as per the manufacturers' recommendations.

In situ hybridization. *In situ* hybridization of endogenous mRNAs and GFP reporter mRNAs was as described³². For the detection of small RNAs, a digoxigenin tail was added to antisense-locked nucleic acid (LNA) oligonucleotides (Exiqon) with the DIG tailing kit (Roche). Tailed LNA oligonucleotides were purified and used for overnight hybridization at 42 °C. All other steps were the same as for mRNAs.

Microinjection. Mature hippocampal neurons⁴³ were microinjected using an AIS2 microinjection system (Cellbiology Trading) attached to a Zeiss Axiovert 200M. Annealed 3'-end labelled (Alexa-546) sense and unmodified antisense strands of miR-134 (IBA) were used at 100 ng μ l⁻¹. *Limk1* RNA was labelled by *in vitro* transcription in the presence of Alexa-488-5' UTP (Molecular Probes) and used at 200 ng μ l⁻¹. Microinjection needles with a tip size between 0.2 and 0.3 μ m were used (P-87, Sutter Instruments) with a holding pressure of 40 hPa and an injection pressure of 80 hPa. Cells were imaged 20 min after injection, and randomly selected images were analysed by three independent observers in a blind manner.

Peptide-mediated delivery. Double-stranded small RNA or 2'-O-methylated DNA oligonucleotides containing a 5' thiol group (80 μ M, IDT) were reduced with TCEP (80 μ M, Sigma) at room temperature for 15 min. Penetratin (80 μ M, Qbiogene) was added and the mixture was incubated at 65 °C followed by 1 h at 37 °C. Coupled oligonucleotides were heated at 65 °C for 15 min before adding to cells at the indicated concentrations for 4 h. Neurons were harvested for western analysis 48 h after transduction.

Image analysis. For spine analysis, neurons were transfected at 8 DIV or 15 DIV with indicated expression plasmids in combination with EGFP and processed for confocal microscopy at 18 DIV. See Supplementary Information for further details on spine analysis.

For Sholl analysis, a series of concentric circles of 10- μ m increments was manually drawn around the cell body, and the number of dendritic intersections at each individual circle was counted. At least ten individual neurons were measured for each experimental condition. To quantify dendritic GFP levels in the local reporter assay, random dendrites were selected based on dsRed staining and plot profiles of the GFP intensity of the same dendrites were derived using ImageJ (NIH). The obtained values were background corrected and normalized to the respective signal in the red channel. At least 20 dendrites per experimental condition of a total of three independent experiments were measured.

Quantitative real-time PCR. Quantitative real-time PCR was performed on a Taq-Man (Perkin Elmer Life Sciences) using the SYBR-green-containing PCR kit (PE Applied Biosystems) as described³².

Preparation of synaptoneurosomes and radio-immunoprecipitation. Synaptoneurosomes were prepared from P15 long-Evans rat pups (Charles River) as described³². For radio-immunoprecipitation, a mouse monoclonal anti-Limk1 antibody (Pharmingen) was used.

Immunocytochemistry. Hippocampal neurons (18 DIV) were immunostained as described³², using a mouse monoclonal anti-Limk1 (Pharmingen) or a rabbit anti-synapsin (Chemicon) antibody as primary antibody.

Luciferase assay. Cortical neurons were transfected at 4 DIV or 12 DIV, and luciferase assays were performed 2 days later with the Dual-Luciferase Reporter Assay System (Promega).

Received 15 August; accepted 25 October 2005.

- West, A. E., Griffith, E. C. & Greenberg, M. E. Regulation of transcription factors by neuronal activity. *Nature Rev. Neurosci.* **3**, 921–931 (2002).
- Kelleher, R. J. III, Govindarajan, A. & Tonegawa, S. Translational regulatory mechanisms in persistent forms of synaptic plasticity. *Neuron* **44**, 59–73 (2004).
- Bartel, D. P. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**, 281–297 (2004).
- He, L. & Hannon, G. J. MicroRNAs: small RNAs with a big role in gene regulation. *Nature Rev. Genet.* **5**, 522–531 (2004).
- Ambros, V. The functions of animal microRNAs. *Nature* **431**, 350–355 (2004).
- Chen, C. Z., Li, L., Lodish, H. F. & Bartel, D. P. MicroRNAs modulate hematopoietic lineage differentiation. *Science* **303**, 83–86 (2004).
- Brennecke, J., Hipfner, D. R., Stark, A., Russell, R. B. & Cohen, S. M. *bantam* encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene *hid* in *Drosophila*. *Cell* **113**, 25–36 (2003).
- Poy, M. N. et al. A pancreatic islet-specific microRNA regulates insulin secretion. *Nature* **432**, 226–230 (2004).
- Chang, S., Johnston, R. J. Jr, Frokjaer-Jensen, C., Lockery, S. & Hobert, O. MicroRNAs act sequentially and asymmetrically to control chemosensory laterality in the nematode. *Nature* **430**, 785–789 (2004).
- Johnston, R. J. & Hobert, O. A microRNA controlling left/right neuronal asymmetry in *Caenorhabditis elegans*. *Nature* **426**, 845–849 (2003).
- Kim, J. et al. Identification of many microRNAs that copurify with polyribosomes in mammalian neurons. *Proc. Natl Acad. Sci. USA* **101**, 360–365 (2004).

12. Krichevsky, A. M., King, K. S., Donahue, C. P., Khrapko, K. & Kosik, K. S. A microRNA array reveals extensive regulation of microRNAs during brain development. *RNA* **9**, 1274–1281 (2003).
13. Lagos-Quintana, M. *et al.* Identification of tissue-specific microRNAs from mouse. *Curr. Biol.* **12**, 735–739 (2002).
14. Giraldez, A. J. *et al.* MicroRNAs regulate brain morphogenesis in zebrafish. *Science* **308**, 833–838 (2005).
15. Miska, E. A. *et al.* Microarray analysis of microRNA expression in the developing mammalian brain. *Genome Biol.* **5**, R68 (2004).
16. Sempere, L. F. *et al.* Expression profiling of mammalian microRNAs uncovers a subset of brain-expressed microRNAs with possible roles in murine and human neuronal differentiation. *Genome Biol.* **5**, R13 (2004).
17. Kiebler, M. A. & DesGroseillers, L. Molecular insights into mRNA transport and local translation in the mammalian nervous system. *Neuron* **25**, 19–28 (2000).
18. Steward, O. & Schuman, E. M. Protein synthesis at synaptic sites on dendrites. *Annu. Rev. Neurosci.* **24**, 299–325 (2001).
19. Eberwine, J., Miyashiro, K., Kacharmina, J. E. & Job, C. Local translation of classes of mRNAs that are targeted to neuronal dendrites. *Proc. Natl Acad. Sci. USA* **98**, 7080–7085 (2001).
20. Campbell, D. S. & Holt, C. E. Chemotropic responses of retinal growth cones mediated by rapid local protein synthesis and degradation. *Neuron* **32**, 1013–1026 (2001).
21. Kang, H. & Schuman, E. M. A requirement for local protein synthesis in neurotrophin-induced hippocampal synaptic plasticity. *Science* **273**, 1402–1406 (1996).
22. Zhang, X. & Poo, M. Localized synaptic potentiation by BDNF requires local protein synthesis in the developing axon. *Neuron* **36**, 675–688 (2002).
23. Rao, A. & Steward, O. Evidence that protein constituents of postsynaptic membrane specializations are locally synthesized: analysis of proteins synthesized within synaptosomes. *J. Neurosci.* **11**, 2881–2895 (1991).
24. Bonhoeffer, T. & Yuste, R. Spine motility. Phenomenology, mechanisms, and function. *Neuron* **35**, 1019–1027 (2002).
25. Hering, H. & Sheng, M. Dendritic spines: structure, dynamics and regulation. *Nature Rev. Neurosci.* **2**, 880–888 (2001).
26. Matsuzaki, M., Honkura, N., Ellis-Davies, G. C. & Kasai, H. Structural basis of long-term potentiation in single dendritic spines. *Nature* **429**, 761–766 (2004).
27. Nagerl, U. V., Eberhorn, N., Cambridge, S. B. & Bonhoeffer, T. Bidirectional activity-dependent morphological plasticity in hippocampal neurons. *Neuron* **44**, 759–767 (2004).
28. Zito, K., Knott, G., Shepherd, G. M., Shenolikar, S. & Svoboda, K. Induction of spine growth and synapse formation by regulation of the spine actin cytoskeleton. *Neuron* **44**, 321–334 (2004).
29. Meister, G., Landthaler, M., Dorsett, Y. & Tuschl, T. Sequence-specific inhibition of microRNA- and siRNA-induced RNA silencing. *RNA* **10**, 544–550 (2004).
30. Hutvagner, G., Simard, M. J., Mello, C. C. & Zamore, P. D. Sequence-specific inhibition of small RNA function. *PLoS Biol.* **2**, E98 (2004).
31. Mansfield, J. H. *et al.* MicroRNA-responsive ‘sensor’ transgenes uncover Hox-like and other developmentally regulated patterns of vertebrate microRNA expression. *Nature Genet.* **36**, 1079–1083 (2004).
32. Schratt, G. M., Nigh, E. A., Chen, W. G., Hu, L. & Greenberg, M. E. BDNF regulates the translation of a select group of mRNAs by a mammalian target of rapamycin–phosphatidylinositol 3-kinase-dependent pathway during neuronal development. *J. Neurosci.* **24**, 9366–9377 (2004).
33. Ji, Y., Pang, P. T., Feng, L. & Lu, B. Cyclic AMP controls BDNF-induced TrkB phosphorylation and dendritic spine formation in mature hippocampal neurons. *Nature Neurosci.* **8**, 164–172 (2005).
34. Bamburg, J. R. Proteins of the ADF/cofilin family: essential regulators of actin dynamics. *Annu. Rev. Cell Dev. Biol.* **15**, 185–230 (1999).
35. Meng, Y. *et al.* Abnormal spine morphology and enhanced LTP in LIMK-1 knockout mice. *Neuron* **35**, 121–133 (2002).
36. Aakalu, G., Smith, W. B., Nguyen, N., Jiang, C. & Schuman, E. M. Dynamic visualization of local protein synthesis in hippocampal neurons. *Neuron* **30**, 489–502 (2001).
37. Takei, N., Kawamura, M., Hara, K., Yonezawa, K. & Nawa, H. Brain-derived neurotrophic factor enhances neuronal translation by activating multiple initiation processes: comparison with the effects of insulin. *J. Biol. Chem.* **276**, 42818–42825 (2001).
38. Rusinov, V., Baev, V., Minkov, I. N. & Tabler, M. MicroInspector: a web tool for detection of miRNA binding sites in an RNA sequence. *Nucleic Acids Res.* **33**, W696–W700 (2005).
39. John, B. *et al.* Human microRNA targets. *PLoS Biol.* **2**, e363 (2004).
40. McAllister, A. K., Katz, L. C. & Lo, D. C. Neurotrophins and synaptic plasticity. *Annu. Rev. Neurosci.* **22**, 295–318 (1999).
41. Lu, B. BDNF and activity-dependent synaptic modulation. *Learn. Mem.* **10**, 86–98 (2003).
42. Martin, K. C. & Kosik, K. S. Synaptic tagging—who’s it? *Nature Rev. Neurosci.* **3**, 813–820 (2002).
43. Goetze, B., Grunewald, B., Kiebler, M. A. & Macchi, P. Coupling the iron-responsive element to GFP—an inducible system to study translation in a single living cell. *Sci. STKE* **2003**, PL12 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank D. Bartel for providing the *lin-41* reporter constructs and northern blot protocols, K. Mizuno for the rat *Limk1* cDNA, G. Corfas for initial help with ISH, Y. Lin and A. West for reagents, J. Bikoff, E. Griffith, E. Hong, S. Paradis and B. Sabatini for critically reading the manuscript, and all the members of the Greenberg laboratory for support and discussion. This work was supported by grants from the NINDS and NICHD (M.E.G.), HFSP (G.S. and M.K.), the Charles Hood Foundation (G.S.), the Hertie Foundation (M.K.), the Schram-Stiftung (M.K.) and a Boehringer Ingelheim Fonds fellowship (F.T.). M.E.G. acknowledges the generous support of the F. M. Kirby Foundation to the Neurobiology Program of Children’s Hospital.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.E.G. (Michael.Greenberg@childrens.harvard.edu).

ARTICLES

The RNA-binding protein FCA is an abscisic acid receptor

Fawzi A. Razem¹, Ashraf El-Kereamy¹, Suzanne R. Abrams² & Robert D. Hill¹

The phytohormone abscisic acid (ABA) regulates various physiological processes in plants. The molecular mechanisms by which this is achieved are not fully understood. Genetic approaches have characterized several downstream components of ABA signalling, but a receptor for ABA has remained elusive. Although studies indicate that several ABA response genes encode RNA-binding or RNA-processing proteins, none has been found to be functional in binding ABA. Here we show that FCA, an RNA-binding protein involved in flowering, binds ABA with high affinity in an interaction that is stereospecific and follows receptor kinetics. The interaction between FCA and ABA has molecular effects on downstream events in the autonomous floral pathway and, consequently, on the ability of the plant to undergo transition to flowering. We further show that ABA binding exerts a direct control on the FCA-mediated processing of precursor messenger RNA. Our results indicate that FCA is an ABA receptor involved in RNA metabolism and in controlling flowering time.

Abscicic acid is a sesquiterpenoid hormone that is recruited by plants as an internal signal to mediate stress responses related to water availability, such as drought-induced stomatal closure^{1–3}. In addition, ABA regulates agronomically important processes during different stages of plant development, including seed drought tolerance, dormancy and storage protein synthesis^{1,3}. Studies on ABA-sensitive mutations link RNA metabolism to ABA signal transduction through the modulation of RNA processing, splicing, stability and degradation^{4–7}. Although several factors involved in the biosynthesis and signalling of ABA have been characterized, a complete itinerary of signalling network components is missing and the identity of ABA receptors remains elusive^{1,2}. ABA signalling seems to be particularly complex. Despite the existence of common elements, none of the genetically defined components is required for all responses to ABA¹. Instead, there is substantial evidence for many redundant mechanisms of ABA perception and signalling. The ABA perception mechanism in the guard cells of *Vicia faba*, for example, has been suggested to be distinct from that occurring in the mesophyll cells⁸. Furthermore, studies using optically pure ABA analogues have suggested the existence of multiple perception sites for ABA with different stereochemical requirements to ABA responses^{1,3,9–11}.

To identify ABA receptors, we screened a barley complementary DNA expression library with anti-idiotypic antibodies (AB2) and isolated a protein, ABAP1 that bound ABA *in vitro*¹². The deduced amino acid sequence of ABAP1 was similar to that of FCA, an *Arabidopsis* protein that regulates flowering time¹³. FCA is a nuclear RNA-binding protein, specific to plants, that promotes flowering by preventing the accumulation of mRNA encoding FLC, a MADS box transcription factor that is a potent repressor of the floral transition^{14–17}. FCA function requires a second protein, the RNA 3'-end processing factor FY, which binds to its tryptophan–tryptophan (WW) protein interaction domain¹⁸. FCA autoregulates its expression by promoting premature cleavage and polyadenylation in intron 3 of its own precursor mRNA (pre-mRNA)^{19,20}. This function of FCA also requires interaction with FY¹⁸.

FCA binds ABA *in vitro*

We first tested whether *Arabidopsis* FCA protein could bind ABA and found that indeed it could. ABA binding to purified recombinant FCA protein increased linearly with increasing concentrations of FCA (Fig. 1a) in a pH-dependent manner that was specific to the wild-type protein (Supplementary Fig. 1a, b). Stereospecificity of FCA binding to the physiologically active (+)-ABA was tested in

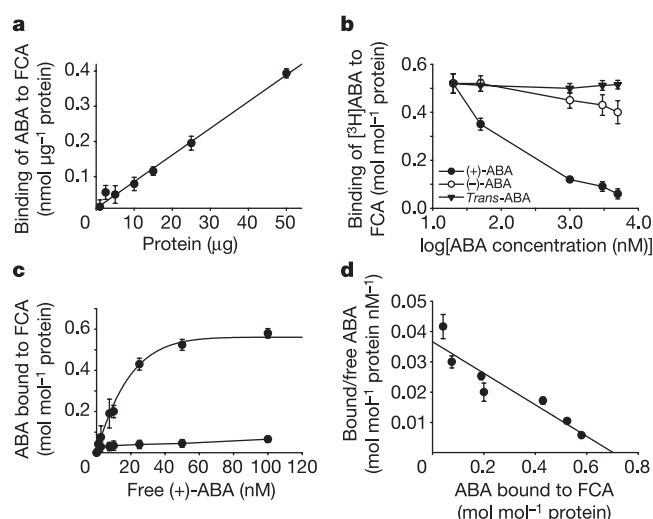


Figure 1 | Binding of [³H](+)-ABA to purified recombinant FCA protein. **a**, Binding of [³H](+)-ABA increases with increasing FCA. **b**, FCA binding of ABA is stereospecific. Only (+)-ABA competes for the binding site. **c**, ABA binding to FCA follows saturation kinetics. Specific binding (top curve) represents the difference between total and nonspecific binding (bottom curve). **d**, Scatchard analysis of ABA binding. Points fitted a linear relationship with the square of the correlation coefficient, $r^2 = 0.88$ ($r^2 = 0.93$ excluding the first ABA concentration). Maximum binding was $0.72 \text{ mol mol}^{-1} \text{ protein}$, with $K_d = 19 \text{ nM}$. Each point is the mean $\pm$ s.d. of nine values.

¹Department of Plant Science, University of Manitoba, Winnipeg, Manitoba R3T 2N2, Canada. ²Plant Biotechnology Institute, National Research Council of Canada, Saskatoon, Saskatchewan S7N 0W9, Canada.

competition assays (Fig. 1b). *Trans*- and (–)-ABA analogues did not compete for the FCA-binding site, at least at concentrations below 3 μM (Fig. 1b).

Specificity of binding was also confirmed using a biotin-tagged (+)-ABA column. FCA eluted from the column with (+)-ABA, but not with (–)-ABA (Supplementary Fig. 2). Specific binding of (+)-ABA to purified FCA could be saturated with increasing amounts of ABA and nonspecific binding was linear and always less than 11% of the total binding (Fig. 1c). Scatchard plot analysis showed a linear relationship (Fig. 1d). FCA bound ABA at a ratio of roughly 0.72 mol of ABA per mol of protein (Fig. 1d), suggestive of a single binding site for ABA. The equilibrium dissociation constant (K_d) for the FCA–ABA complex was 19 nM. Thus, the ABA-binding kinetics of purified recombinant FCA protein meet the basic characteristics of an ABA-binding protein.

ABA disrupts the FCA–FY interaction *in vitro*

We next examined how the binding of ABA to FCA might affect the function of FCA. Because the autoregulatory activity and *FLC* repression function of FCA depend on its association with the 3' RNA-processing factor FY, we tested whether ABA binding influenced this protein–protein interaction. Recombinant FCA protein, translationally fused to glutathione *S*-transferase (GST), was purified from *Escherichia coli*. The binding of FCA to *in vitro* translated FY labelled with [^{35}S]methionine was assessed through pull-down assays using glutathione agarose.

We found that the interaction of FY with FCA was markedly inhibited by (+)-ABA (Fig. 2a). Consistent with the stereospecificity observed in ABA binding to FCA, however, (–)-ABA had no effect on the interaction of FCA with FY (Fig. 2a). Inhibition of the FCA–FY interaction by ABA was dependent on concentration, with less than 10% of the maximum FCA–FY interaction detected in the presence of 1 μM ABA (Fig. 2b). FCA pre-complexed with FY was then tested for its ability to bind ABA. Initially, its ABA-binding activity was low, but it increased significantly ($P < 0.05$, *t* test) after 45 min of incubation and paralleled a decrease in the association of FY with FCA (Supplementary Fig. 3). This finding indicated that either ABA could disrupt FCA–FY complexes or, in the course of FCA–FY

association and dissociation, ABA could prevent FCA–FY reassociation.

We tested whether two conserved domains, the RNA recognition motif (RRM) at the amino terminus and the WW domain at the C terminus of FCA, might have any role in ABA binding. ABA-binding assays using crude lysate prepared from *E. coli* expressing either the N terminus (FCA-RRM) or the C terminus (FCA-WW) of FCA showed that ABA binding occurred in the C terminus containing the WW domain (Fig. 2c). FY binds through the WW protein interaction domain of FCA, and mutation of the second signature tryptophan residue of the FCA-WW domain to phenylalanine (FCA-WF) abolishes the FCA–FY interaction¹⁸. We found that ABA bound FCA wild-type and the FCA-WF mutant protein equally well, however, indicating that ABA and FY do not bind FCA at precisely the same site (Fig. 2d). Thus, ABA has the potential to inhibit FCA activity by disrupting its association with the RNA-processing factor FY, with the ABA-binding site being located nearer the C terminus of FCA.

ABA treatment phenocopies disruption of FCA function *in vivo*

Although proteins that bind ABA have been identified, no evidence has been presented to link them to physiological effects of ABA *in vivo*. To assess the significance of ABA disrupting the FCA–FY interaction *in vitro*, we studied the effect of ABA on two activities of FCA *in vivo*: FCA autoregulation and flowering time control. FCA autoregulates its expression by promoting premature cleavage and polyadenylation in its own pre-mRNA^{19,20}. Truncated *FCA β* transcripts, cleaved and polyadenylated in intron 3, and fully spliced *FCA γ* transcripts, cleaved and polyadenylated in the conventional 3' untranslated region, can be detected in wild-type plants²⁰. In genetic backgrounds lacking either functional FCA or FY protein, the ratio of these transcripts to one another changes as compared with wild type: the amount of full-length *FCA γ* transcript increases, and there is a reciprocal decrease in the truncated, prematurely cleaved and polyadenylated transcript *FCA β* ²⁰. We treated wild-type plants with ABA and assessed its effects on alternative processing of *FCA* pre-mRNA. RNA blot analysis showed that ABA induced a shift in the ratio of *FCA* isoforms (Fig. 3a) similar to that seen in loss-of-function *fca* or *fy* genotypes, with less *FCA β* and more *FCA γ* transcripts in ABA-treated plants than in untreated plants. This observation is consistent with ABA acting to disrupt FCA function by preventing the association of FY with FCA.

The autoregulation of *FCA* expression is developmentally controlled. A reporter gene fusion comprising *FCA* promoter and gene sequence up to exon 5 fused to the gene encoding β -glucuronidase (*FCA::to exon5::GUS*) stably expressed in transgenic *Arabidopsis* plants

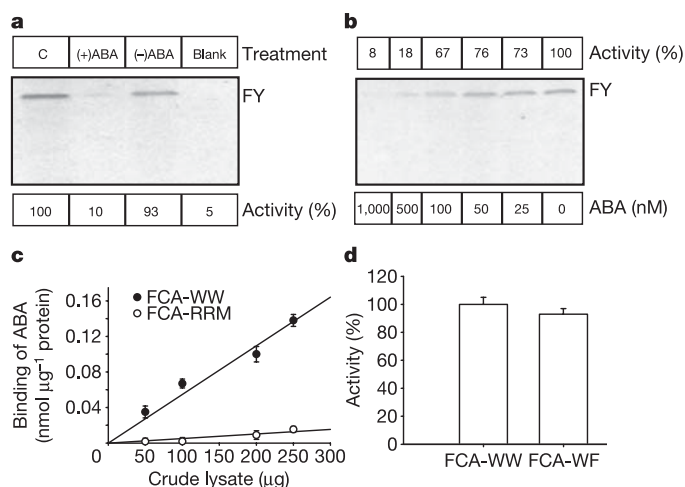


Figure 2 | ABA affects FCA–FY interaction. **a**, Only (+)-ABA (1 μM) inhibits FCA–FY interaction in GST pull-down assays. Activity is expressed as a percentage of the control (C) value (2.5×10^3 d.p.m.). **b**, ABA inhibits FCA–FY interaction in a dose-dependent manner. ^{35}S activity in the FY band, expressed as a percentage of the radioactivity in the absence of ABA (2.1×10^3 d.p.m.), is indicated at the top. **c**, FCA binds ABA only at the C-terminal end containing the WW domain. **d**, FCA does not bind ABA at the FY-binding site in the WW domain. Each point (in **c**, **d**) is the mean $\pm$ s.d. of three values.

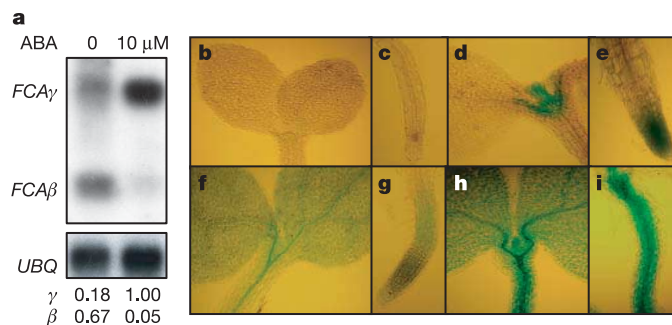


Figure 3 | ABA affects FCA autoregulation. **a**, Interference of FCA autoregulation by ABA shifts the expression ratio between the β and γ transcripts of *FCA*. Values below the blot represent the ratio of *FCA β* to ubiquitin (*UBQ*), with the maximum value set to 1. **b–i**, Histochemical GUS activity in shoots and roots of transgenic plants expressing *P_{FCA}–FCA::to exon5::GUS* (refs 19, 20). Shown are 2-day-old (**b**, **c**, **f**, **g**) and 5-day-old (**d**, **e**, **h**, **i**) plants in the absence (**b–e**) or presence (**f–i**) of ABA.

reports the alternative processing of *FCA* pre-mRNA in different genetic backgrounds^{19,20}. GUS activity is dependent on the excision of *FCA* intron 3 by pre-mRNA splicing and reports on *FCA* γ formation. In wild-type plants expressing *FCA::to exon5::GUS*, increased GUS expression is detected only 4–6 d after germination and is restricted to shoot and root apical regions. In *fca* or *fy* mutant backgrounds defective in *FCA* autoregulation, however, increased GUS expression is detected earlier in development and shows a much wider pattern of expression^{19,20}. Treatment of wild-type seedlings expressing *FCA::to exon5::GUS* with ABA also resulted in the detection of increased GUS expression earlier in development and with a much broader pattern of distribution (Fig. 3b–i and Supplementary Fig. 4). Therefore, treatment of intact *Arabidopsis* seedlings with ABA phenocopied the effects of loss of *FCA* function on the autoregulatory activity of *FCA*. Thus, the *in vivo* effects of ABA are consistent with its effects on *FCA*–*FY* complex formation *in vitro*.

We next tested how ABA affected the function of *FCA* in controlling flowering time. If ABA inhibits the interaction of *FCA* with *FY* *in vivo*, ABA treatment should delay *Arabidopsis* flowering. Consistent with this expectation, the application of ABA delayed the timing of the switch to flowering in wild-type plants, causing an increase in the days to bolting and an increase in rosette leaf number (Fig. 4a and Supplementary Fig. 5). To determine whether this delay was due specifically to ABA disrupting *FCA* function rather than to a pleiotropic effect of ABA that coincidentally affected *FLC* expression,

we treated *fca-1* and *fy-1* mutants with ABA (Fig. 4b). *FCA* comprises one component of the genetically defined autonomous pathway that prevents the accumulation of *FLC* mRNA. Components of this pathway regulate *FLC* through different mechanisms and show some genetic redundancy¹⁵. However, addition of ABA to *fca-1* and *fy-1* plants did not cause any further delay in flowering (Fig. 4b).

To ensure that the ABA effects on wild-type flowering time are specific for (+)-ABA, we treated wild-type plants with two of the non-active ABA analogues, (–)-ABA and *trans*-(+)-ABA and found that, in contrast to (+)-ABA, neither analogue significantly delayed flowering time (Fig. 4a). The inhibitory effect of exogenously applied ABA on flowering was also consistent with early flowering observed in mutants that affect ABA biosynthesis. For example, the *Arabidopsis* ABA-deficient mutant *aba1* flowered early in the absence of ABA, but at a similar time to wild-type after exogenous ABA application (Fig. 4a). This is in agreement with the inhibitory effect of ABA on floral initiation that has been reported in several plant species after exogenous application of ABA^{21–23} and has been also suggested by others (see ref. 24 for citations). Furthermore, more than 20-fold lower endogenous ABA levels are present at early floral initiation than at vegetative growth in *Polianthes tuberosa*²³. Consistent with ABA disrupting the activity of *FCA*, late flowering after ABA treatment was associated with relatively increased levels of mRNA encoding the floral repressor, *FLC*, in wild-type and *aba1* plants (Fig. 4c–e) but not in *fca-1* mutants (Fig. 4f). The increased

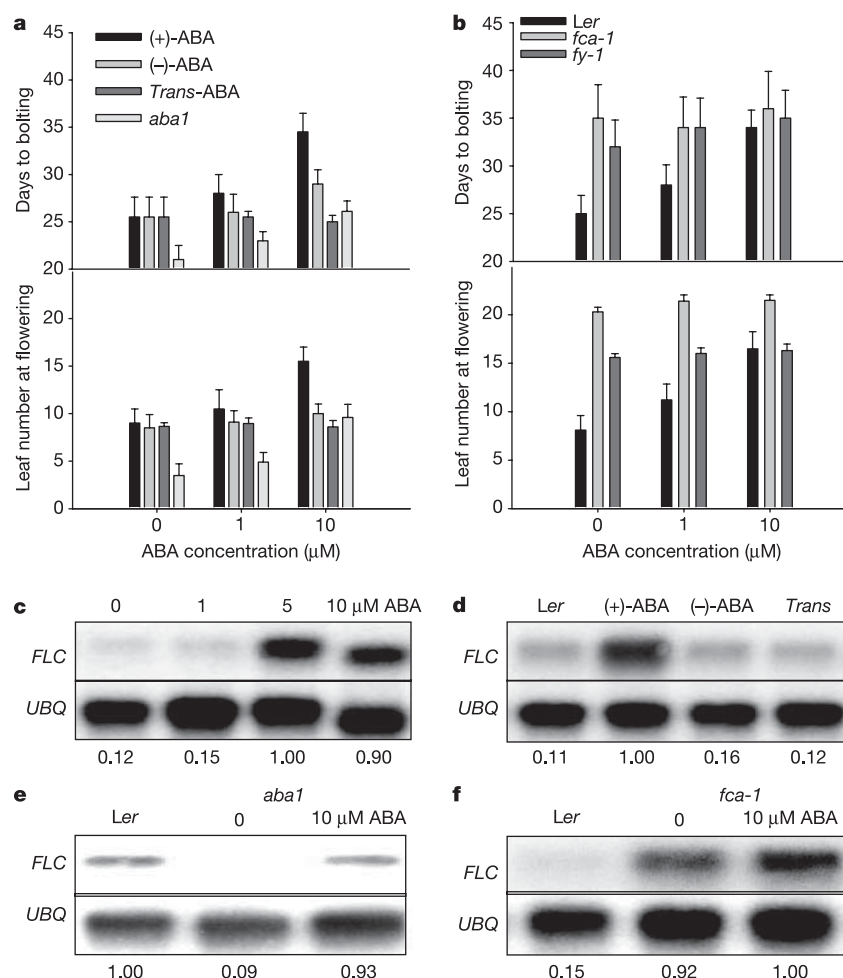


Figure 4 | ABA affects flowering. **a**, Days to bolting and rosette leaf number at flowering after treatment of wild-type plants with (+)-ABA, (–)-ABA and *trans*-ABA, and treatment of *aba1* plants with (+)-ABA. **b**, Flowering in *fca-1* and *fy-1* mutants after (+)-ABA treatment. Error bars in **a**, **b** represent the s.d. **c–f**, (+)-ABA increases *FLC* mRNA in wild-type (**c**, **d**) and *aba1*

(**e**), but not in *fca-1* (**f**) or *fy-1* mutants (not shown). Plants were treated every other day with 1, 5 or 10 μ M (+)-ABA (**c**), or with 10 μ M (+)-ABA (**d–f**) for 3 weeks. Values below the blot represent the ratio of *FLC* to ubiquitin (*UBQ*), with the maximum value set to 1.

level of *FLC* mRNA was associated only with application of the physiologically active (+)-ABA (Fig. 4d).

Taken together, therefore, the *in vivo* effects of ABA mimic the stereospecificity of our *in vitro* analysis and are consistent with ABA disrupting FCA function by antagonizing the interaction of FCA and FY, with a functional consequence on flowering time control.

FCA is involved in a distinct ABA response

Because ABA affects several physiological processes in plants such as seed germination, stomatal aperture and root development, we tested whether ABA, through FCA, could affect some of these responses. First, we examined whether FCA has roles in seed germination and stomatal response. In the presence of ABA, none of the *fca-1* seeds germinated and the aperture of *fca-1* guard cells significantly ($P < 0.05$) decreased after ABA application (Fig. 5a), in a fashion similar to the effects of ABA on wild-type plants (Fig. 5a). Thus, FCA is not required for seed germination or stomatal response.

We then tested whether FCA has a role in lateral root formation. *fca-1* mutants possess fewer lateral root numbers per unit area¹⁹. ABA has been shown to affect lateral root formation in *Arabidopsis*, probably through a signalling pathway that differs from any of the known ABA response genes such as *ABI1* and *ABI2* (ref. 25). We counted lateral roots in 2-week-old *fca-1* mutants grown in the presence and absence of ABA (Fig. 5b). In the presence of ABA, the number of lateral roots decreased only slightly in *fca-1*, but decreased considerably in wild-type plants (Fig. 5b). Thus, FCA may have a role in ABA-inhibited lateral root formation, but further work is required to understand the mechanisms involved.

To examine more closely this redundancy in ABA signalling, we tested whether ABA-insensitive *abi-1* and *abi-2* mutations could disrupt FCA function. Application of ABA caused a delay in flowering time (Fig. 5c and data not shown for *abi-1*) similar to that observed in the wild type, whereas ABA had no effect on *fca-1* mutant plants (Fig. 4), indicating FCA does not function through either of the pathways involving the ABA-insensitive genes. We then examined FCA mRNA processing in *abi-2* by testing whether treatment of

these seedlings with ABA could affect the ratio of *FCAβ* to *FCAγ* transcripts (Fig. 5d). RNA blot analysis showed a shift in the ratio of FCA isoforms in *abi-2* (Fig. 5d) similar to that observed in the wild-type after ABA treatment (Fig. 3a). Thus, FCA and ABI proteins are involved in distinct ABA responses. These data also provide further evidence for the complexity and multitude of receptor and signalling mechanisms in plants.

Discussion

Here we have provided evidence that FCA is an ABA receptor involved in a physiological process that is distinct from other known ABA responses. Our results provide compelling evidence supporting the existence of several perception sites for ABA in plants. In addition, results presented here and in published reports^{26,27} show that hormone signalling could be unconventional in its nature and may involve only a direct protein–protein interaction that is positively or negatively influenced by hormone binding. Our data establish a direct functional link between ABA signalling and flowering time control through the action of ABA on the interaction of FCA with FY and the 3' end RNA-processing machinery. The role of FCA in promoting flowering is accounted for by its regulation of *FLC*^{16,28–30}, however, FCA is more widely conserved in other plant species than is *FLC*. Exploring the effect FCA has on ABA responses in other species may identify other functions of FCA that account for its evolutionary conservation. It will be interesting to determine which other functions of ABA depend on FCA.

Our data also establish a direct link between ABA signalling and RNA processing. Of the elements that affect ABA signalling, factors that affect RNA processing are relatively prominent^{7,31,32}. Mutations in the Cap-binding protein, Sm-like protein and HYL1 (a double-stranded RNA-binding protein that affects micro-RNA biogenesis) all show ABA hypersensitive responses^{4–6}. In addition, ABA signalling in stomata guard cells involves regulation of the activity and sub-nuclear localization of the hnRNP D-like RNA-binding protein AKIP1 (ref. 32). These multiple connections between ABA signalling and RNA processing indicate either that ABA signalling is used much more in the posttranscriptional regulation of gene expression than are other plant hormone signalling cascades or that these RNA connections remain to be uncovered in other plant hormone signalling pathways.

To our knowledge, RNA-binding proteins, such key regulators of gene expression, have not been found to be direct targets of hormone action; our observations therefore define a new class of receptor. The *Arabidopsis* genome encodes many RNA-binding proteins, including 198 that possess the most common RNA-binding protein domain found in nature, RRM³¹. About half of these RRM-containing proteins are uncharacterized and plant-specific. Plant development and stress responses depend on the continuous integration of numerous signals, many of them small molecules. With so many uncharacterized RNA-binding proteins in plants and other organisms, it will be interesting to determine whether direct hormone-mediated mechanisms of posttranscriptional gene regulation are widespread.

METHODS

Materials, growth conditions and determination of GUS activity. Authentic ABA analogues were provided by the National Research Council of Canada, Saskatoon, Saskatchewan. All *Arabidopsis thaliana* genotypes were on a Landsberg *erecta* background. Wild-type (Ler), *fca-1*, *fy-1*, *abi-1*, *abi-2* and *aba1* seeds were obtained from the Arabidopsis Biological Resource Center. Plant growth conditions have been described¹⁹. For RNA isolation, *Arabidopsis* seeds were stratified for 3 d at 4 °C, sown in soil under 16 h of light, and sprayed every other day with different concentrations of ABA, starting 4 d after germination for a period of 3 weeks. For flowering time, with each individual measurement: 158 Ler plants were used to study the effects of (+)-ABA; 48 to study effects of (–)- and *trans*-ABA; 38 plants to study effects of (+)-ABA on *fca*, *fy* and *aba1* mutants; and 33 to study (+)-ABA effects on *abi-2* mutants. Growth conditions and ABA spray were done as described above. Lateral roots of 21 Ler plants and 26 *fca-1* mutant plants were measured after ABA application as described²⁵. The

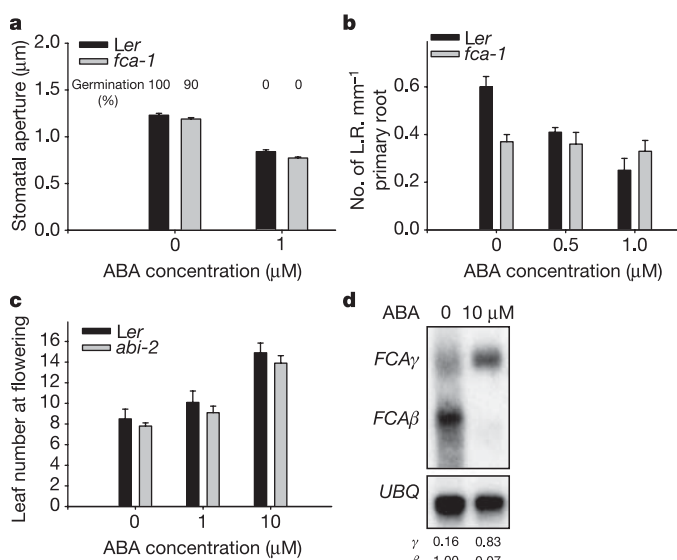


Figure 5 | FCA is involved in a distinct ABA response. **a**, ABA inhibits seed germination (top numbers) and decreases stomatal aperture in *fca-1* and Ler. **b**, *fca-1* roots do not respond to ABA. The difference in lateral root (L.R.) numbers between treatments was only significant ($P < 0.05$) in Ler. **c**, Leaf number at flowering after ABA treatment in Ler and *abi-2*. Error bars in **a–c** represent the s.d. **d**, ABA shifts the expression ratio between *FCAβ* and *FCAγ* transcripts. Plants treated every other day with 10 μM (+)-ABA for 3 weeks. Values below the blot represent the ratio of FCA to ubiquitin (UBQ), with the maximum value set to 1.

aperture of 50 stomata and the germination of 50 seeds per treatment were determined in *Ler* and *fca-1* mutants in the absence and presence of ABA. For GUS staining, transgenic plants were grown in Petri dishes (under 16 h of light) on GM medium¹⁹ and sprayed with 10 μ M ABA 12 h before harvest. Seedlings were transferred to 1 mM 5-bromo-4-chloro-3-indolyl- β -D-glucuronide and processed as described³³.

GST-binding assays of FCA–FY interaction. All *in vitro* translation and GST pull-down assays were done in accordance with the Promega Technical Manual No. 249. FY protein was synthesized from a plasmid template and labelled with [³⁵S]methionine using the T7 TNT coupled Transcription/Translation System (Promega). The protein-binding or protein interaction reaction was carried out for 90 min at 4 °C with continuous gentle mixing. To test the effect of ABA on FCA–FY interaction, GST–FCA was incubated in the presence or absence of 1 μ M (–) or (+)-ABA except for the concentration-dependent study, which used (+)-ABA (Fig. 2b). FY translated product was added to the incubation mixture 30 min after ABA addition and the reaction was allowed to proceed for an additional 90 min. The mixture was centrifuged and the pellet was washed, resuspended in 20 μ l of SDS–PAGE loading buffer containing 10 mM glutathione, boiled for 5 min and loaded on SDS–PAGE. After electrophoresis, the gel was dried and the labelled proteins were detected in accordance with the Promega Technical Manual No. 249.

Purification of RNA and northern blots. RNA was isolated by using guanidine and a standard method. mRNA was isolated with a PolyATtract mRNA Isolation System (Promega). For northern blots, RNA was separated by electrophoresis on 1.2% agarose containing 8% formaldehyde and transferred to Hybond N membrane (Amersham). For FCA analysis, we used RT–PCR to prepare an FCA 5' leader probe using primers as described²⁰. Northern blot analysis was carried out on ~0.9 μ g of mRNA that had been blotted with the FCA 5' leader probe²⁰ and then rehybridized with UBQ10. *FLC* was analysed as described above (using ~0.43 μ g of mRNA) with a probe made from the larger fragment of an *FLC* clone (a gift from R. Amasino) digested with *Eco*RI and *Sph*I. To normalize loading, membranes were stripped in 0.1% SDS and rehybridized with a probe for the ubiquitin (UBQ10) coding region. All probes were labelled with [α -³²P]dCTP. Values below each blot represent the ratio of target gene to ubiquitin, with the maximum value set to 1.

ABA-binding assays. Crude lysate and purified FCA protein were used to determine the ABA-binding activity as described¹². FCA–RRM was purified from a clone containing the 5' end of *FCA* γ , including the two RRM, in pGEX-6P-1 digested with *Eco*RI–*Xho*I into which an *Eco*RI–*Xho*I-digested PCR fragment had been ligated. The FCA–WW was purified from a clone containing the 3' end of *FCA* γ , including the WW interaction domain, in pGEX-6P-1 as above, except that digestion was with *Sal*I–*Eco*RI.

For competitive assays, the ABA analogues (–)-ABA and *trans*-ABA were added at the same time as [³H](+)-ABA at different concentrations (20–5,000 nM). Specific binding was calculated by taking the difference for assays with only [³H](+)-ABA (total binding) and assays that also contained 5 μ M (+)-ABA added at the same time as [³H](+)-ABA (nonspecific binding). Purification of FCA protein on biotin-tagged (+)-ABA column³⁴ was done as described in the figure legends. Binding is represented as the number of moles of [³H](+)-ABA per mole of FCA protein. To determine FY dissociation from the FCA–FY complex in the presence of 1 μ M ABA, the interaction reaction was stopped after ABA addition at the time points shown in Supplementary Fig. 3, and the mixture was washed and resuspended in 20 μ l of 10 mM GSH. Dual activity for ³⁵S and ³H was counted simultaneously on a scintillation counter. Activity is expressed as a percentage of the control value (in the absence of [³H](+)-ABA for FY; in the absence of ³⁵S-FY for ABA). The highest d.p.m. for each control was taken as 100% and is given in the legend to Supplementary Fig. 3.

Received 15 August; accepted 24 October 2005.

1. Finkelstein, R. R., Gampala, S. S. L. & Rock, C. D. Abscisic acid signalling in seeds and seedlings. *Plant Cell* **14**, S15–S45 (2002).
2. Himmelbach, A., Yang, Y. & Grill, E. Relay and control of abscisic acid signalling. *Curr. Opin. Plant Biol.* **6**, 470–479 (2003).
3. Leung, J. & Giraudat, J. Abscisic acid signal transduction. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **49**, 199–222 (1998).
4. Lu, C. & Fedoroff, N. A mutation in the *Arabidopsis* *HYL1* gene encoding a dsRNA binding protein affects responses to abscisic acid, auxin, and cytokinin. *Plant Cell* **12**, 2351–2365 (2000).
5. Hugouvieux, V., Kwak, J. & Schroeder, J. A mRNA cap binding protein, ABH1, modulates early abscisic acid signal transduction in *Arabidopsis*. *Cell* **106**, 477–487 (2001).
6. Xiong, L. *et al.* Modulation of abscisic acid signal transduction and biosynthesis by an Sm-like protein in *Arabidopsis*. *Dev. Cell* **1**, 771–781 (2001).
7. Kuhn, J. M. & Schroeder, J. I. Impacts of altered RNA metabolism on abscisic

- acid signalling. *Curr. Opin. Plant Biol.* **6**, 463–469 (2003).
8. Sutton, F., Paul, S. S., Wang, X.-Q. & Assmann, S. M. Distinct abscisic acid signalling pathways for modulation of guard cells versus mesophyll cell potassium channels revealed by expression studies in *Xenopus laevis* oocytes. *Plant Physiol.* **124**, 223–230 (2000).
9. Gusta, L. V., Ewan, B., Reaney, J. T. & Abrams, S. R. The effect of abscisic acid and abscisic acid metabolites on the germination of cress seed. *Can. J. Bot.* **70**, 1550–1555 (1992).
10. Hill, R. D. *et al.* Abscisic acid structure–activity relationships in barley aleurone layers and protoplasts. *Plant Physiol.* **108**, 573–579 (1995).
11. Hays, D. B., Rose, P., Abrams, S. R. & Moloney, M. M. Biological activity of optically pure C-1 altered abscisic acid analogs in *Brassica napus* microspore embryos. *J. Plant Growth Reg.* **15**, 5–11 (1996).
12. Razem, F. A., Luo, M., Liu, J.-H., Abrams, S. R. & Hill, R. D. Purification and characterization of a barley aleurone abscisic acid-binding protein. *J. Biol. Chem.* **279**, 9922–9929 (2004).
13. Macknight, R. *et al.* FCA, a gene controlling flowering time in *Arabidopsis*, encodes a protein containing RNA-binding domains. *Cell* **89**, 737–745 (1997).
14. Henderson, I. R. & Dean, C. Control of *Arabidopsis* flowering: the chill before the bloom. *Development* **131**, 3829–3838 (2004).
15. Simpson, G. G. The autonomous pathway: epigenetic and post-transcriptional gene regulation in the control of *Arabidopsis* flowering time. *Curr. Opin. Plant Biol.* **7**, 570–574 (2004).
16. Michaels, S. D. & Amasino, R. M. FLOWERING LOCUS C encodes a novel MADS domain protein that acts as a repressor of flowering. *Plant Cell* **11**, 949–956 (1999).
17. Sheldon, C. C., Rouse, D. T., Finnegan, E. J., Peacock, W. J. & Dennis, E. S. The molecular basis of vernalization: the central role of FLOWERING LOCUS C (*FLC*). *Proc. Natl Acad. Sci. USA* **97**, 3753–3758 (2000).
18. Simpson, G. G., Dijkwel, P. P., Quesada, V., Henderson, I. & Dean, C. FY is an RNA 3' end-processing factor that interacts with FCA to control the *Arabidopsis* floral transition. *Cell* **113**, 777–787 (2003).
19. Macknight, R. *et al.* Functional significance of the alternative transcript processing of the *Arabidopsis* floral promoter FCA. *Plant Cell* **14**, 877–888 (2002).
20. Quesada, V., Macknight, R., Dean, C. & Simpson, G. G. Autoregulation of FCA pre-mRNA processing controls *Arabidopsis* flowering time. *EMBO J.* **22**, 3142–3152 (2003).
21. Wijayanti, L., Fujioka, S., Kobayashi, M. & Sakurai, A. Involvement of abscisic acid and indole-3-acetic acid in the flowering of *Pharbitis nil*. *J. Plant Growth Regul.* **16**, 115–119 (1997).
22. Markarov, A. M. Causes of flowering of long-day potato species under short-day and cold-night conditions. *Russ. J. Plant Physiol.* **49**, 465–469 (2002).
23. Su, W. R., Huang, K. L., Shen, R. S. & Chen, W. S. Abscisic acid affects floral initiation in *Polianthes tuberosa*. *J. Plant Physiol.* **159**, 557–559 (2002).
24. Finkelstein, R. R. & Rock, C. D. In *The Arabidopsis Book* (eds Somerville, C. R. & Meyerowitz, E. M.) 1–48 (Am. Soc. Plant Biologists, Rockville, MD, 2002).
25. De Smet, I. *et al.* An abscisic acid-sensitive checkpoint in lateral root development of *Arabidopsis*. *Plant J.* **33**, 543–555 (2003).
26. Dharmasiri, N., Dharmasiri, S. & Estelle, M. The F-box protein TIR1 is an auxin receptor. *Nature* **435**, 441–445 (2005).
27. Kepinski, S. & Leyser, O. The *Arabidopsis* F-box protein TIR1 is an auxin receptor. *Nature* **435**, 446–451 (2005).
28. Michaels, S. D. & Amasino, R. M. Loss of FLOWERING LOCUS C activity eliminates the late-flowering phenotype of *FRIGIDA* and autonomous pathway mutations but not responsiveness to vernalization. *Plant Cell* **13**, 935–941 (2001).
29. Bastow, R. *et al.* Vernalization requires epigenetic silencing of *FLC* by histone methylation. *Nature* **427**, 164–167 (2004).
30. Sung, S. & Amasino, R. M. Remembering winter: towards a molecular understanding of vernalization. *Annu. Rev. Plant Biol.* **56**, 491–508 (2005).
31. Cheng, Y. & Chen, X. Posttranscriptional control of plant development. *Curr. Opin. Plant Biol.* **7**, 20–25 (2004).
32. Li, J. *et al.* Modulation of an RNA-binding protein by abscisic-acid-activated protein kinase. *Nature* **418**, 793–797 (2002).
33. Jefferson, R. A., Kavanagh, T. A. & Bevan, M. W. GUS fusions: β -glucuronidase as a sensitive and versatile gene fusion marker in higher plants. *EMBO J.* **6**, 3901–3907 (1987).
34. Nyangulu, J. M. *et al.* An affinity probe for isolation of abscisic-acid proteins. *J. Am. Chem. Soc.* **127**, 1662–1664 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank C. Dean and G. Simpson for clones and advice; M. Galka for an ABA affinity column; and K. Baron for help with microscopy. The financial support of the Natural Sciences and Engineering Council of Canada (to R.D.H.) and Genome Canada (to R.D.H. and S.R.A.) is gratefully acknowledged.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to R.D.H. (rob_hill@umanitoba.ca).

A late Miocene dust shower from the break-up of an asteroid in the main belt

Kenneth A. Farley¹, David Vokrouhlický², William F. Bottke³ & David Nesvorný³

Throughout the history of the Solar System, Earth has been bombarded by interplanetary dust particles (IDPs), which are asteroid and comet fragments of diameter $\sim 1\text{--}1,000\ \mu\text{m}$. The IDP flux is believed to be in quasi-steady state: particles created by episodic main belt collisions or cometary fragmentation replace those removed by comminution, dynamical ejection, and planetary or solar impact. Because IDPs are rich in ^3He , seafloor sediment ^3He concentrations provide a unique means of probing the major events that have affected the IDP flux and its source bodies over geological timescales^{1–4}. Here we report that collisional disruption of the $>150\text{-km}$ -diameter asteroid that created the Veritas family $8.3 \pm 0.5\text{ Myr}$ ago⁵ also produced a transient increase in the flux of interplanetary dust-derived ^3He . The increase began at $8.2 \pm 0.1\text{ Myr}$ ago, reached a maximum of ~ 4 times pre-event levels, and dissipated over $\sim 1.5\text{ Myr}$. The terrestrial IDP accretion rate was overwhelmingly dominated by Veritas family fragments during the late Miocene. No other event of this magnitude over the past $\sim 10^8\text{ yr}$ has been deduced from main belt asteroid orbits. One remarkably similar event is present in the ^3He record 35 Myr ago, but its origin by comet shower¹ or asteroid collision⁶ remains uncertain.

After release from their comet or asteroid parent bodies, IDPs spiral towards the Sun under the effects of non-gravitational forces including Poynting–Robertson (P–R) and solar wind drags⁷. P–R drag occurs when dust grains revolving around the Sun absorb solar photons and then reradiate the energy in all directions. At the same time, implantation of solar wind ions enriches IDPs in ^3He . If these particles avoid intense frictional heating during atmospheric entry, they can reach the Earth's surface with their ^3He intact⁸. We have obtained ^3He data on sediments spanning the past 70 Myr (refs 1, 9, 10), with new data in the interval $3\text{--}38\text{ Myr}$ ago reported here. We analysed two pelagic carbonate cores: ODP Site 757 in the Indian Ocean over this entire interval, and Site 926 in the Atlantic Ocean in the late Miocene.

^3He measurements indicate the IDP flux is characterized by a somewhat bumpy continuum punctuated by sharp peaks at 8.2 and 35 Myr ago (Fig. 1). The older of these peaks, in the late Eocene, has been described from a different locality¹. This peak is well above the average of the past 70 Myr and is coincident with the formation of the two largest terrestrial impact craters of the Cenozoic era: Popagai and Chesapeake Bay. The simultaneous increase in the dust and large body flux, and the match between the duration of the dust spike and that predicted for the ejection timescale of long period comets, were taken as evidence for a comet shower, perhaps produced by a close stellar encounter¹. However the composition of impact melt at Popagai crater suggests an L-chondrite impactor⁶, implying that asteroids rather than comets may produce the spikes in IDP flux. Although several other episodes of elevated flux have been hinted at,

none have yet been confirmed. Possible connections between the late Miocene and Eocene events are further discussed in Supplementary Information.

To assess the distribution and temporal evolution of the late Miocene (8.2 Myr ago) peak, we studied the event at higher temporal resolution (Fig. 2) at Site 757 and also at Site 926. A ^3He flux peak beginning 8.2 Myr ago and with nearly identical relative magnitude (factor of ~ 4 above pre-event values) and duration ($\sim 1.5\text{ Myr}$) is apparent at both sites. The only major distinction between the records is that the flux at Site 926 is about three times higher than at Site 757. This probably reflects the effects of sediment focusing, which is known to occur at Site 926 (ref 11). Given the similarity of the ^3He peak at these two sites, and the fact that the peak does not correspond to dramatic changes in sediment composition or sedimentation rate (Supplementary Information), it seems unlikely that it is a sedimentation artefact. Furthermore, at both sites the flux peak corresponds to peaks in ^3He concentration, $^3\text{He}/^4\text{He}$ ratio and $^3\text{He}/\text{non-carbonate fraction}$ (Supplementary Fig. S1). These observations indicate an increase in the IDP flux⁹. Thus we conclude that this ^3He peak, like that in the late Eocene, is a global signature of an IDP-producing astronomical event.

Although the late Miocene and late Eocene ^3He peaks are similar in duration and magnitude (Fig. 1), there is one important difference. Unlike the late Eocene with its two large impact craters that demand an increase in the large body flux coincident with the IDP spike, no late Miocene craters have yet been found. Apparently the late Miocene event was not accompanied by an asteroid or comet shower. This suggests the need for a mechanism capable of increasing the flux of IDPs striking Earth without affecting the flux of larger bodies. A likely candidate is the disruption of the diameter $D > 150\text{ km}$ asteroid that produced the Veritas family, a cluster of fragments on similar orbits at 3.17 AU . The Veritas event was the largest asteroid disruption in the past 10^8 yr (refs 5, 12); resulting collisions still produce as much as 10% of all Solar System near-ecliptic dust^{13–15}. The age of the family, $8.3 \pm 0.5\text{ Myr}$ (ref. 5) was determined by tracking the orbits of Veritas family members backwards in time to their formation (for details, see Supplementary Fig. S4), and coincides with the onset of the late Miocene ^3He spike.

When the parent body of Veritas disrupted, it ejected almost half of its mass in the form of fragments ranging from micrometre-sized dust grains to multi-kilometre asteroids⁵. These bodies then experienced dynamical evolution according to size. The evolution of small fragments ($D = 1\text{--}1,000\ \mu\text{m}$) was dominated by planetary perturbations and non-gravitational forces, which caused them to spiral inwards towards the Sun. In contrast, larger fragments ($D > 1,000\ \mu\text{m}$) were trapped in the main belt unless they could reach a chaotic resonance capable of placing them onto a planet-crossing orbit. The nearest resonances capable of producing an

¹Division of Geological and Planetary Sciences, California Institute of Technology, MS 170-25, Pasadena, California 91125, USA. ²Institute of Astronomy, Charles University, V Holešovičkách 2, 180 00 Prague 8, Czech Republic. ³Department of Space Studies, Southwest Research Institute, 1050 Walnut Street, Suite 400, Boulder, Colorado 80302, USA.

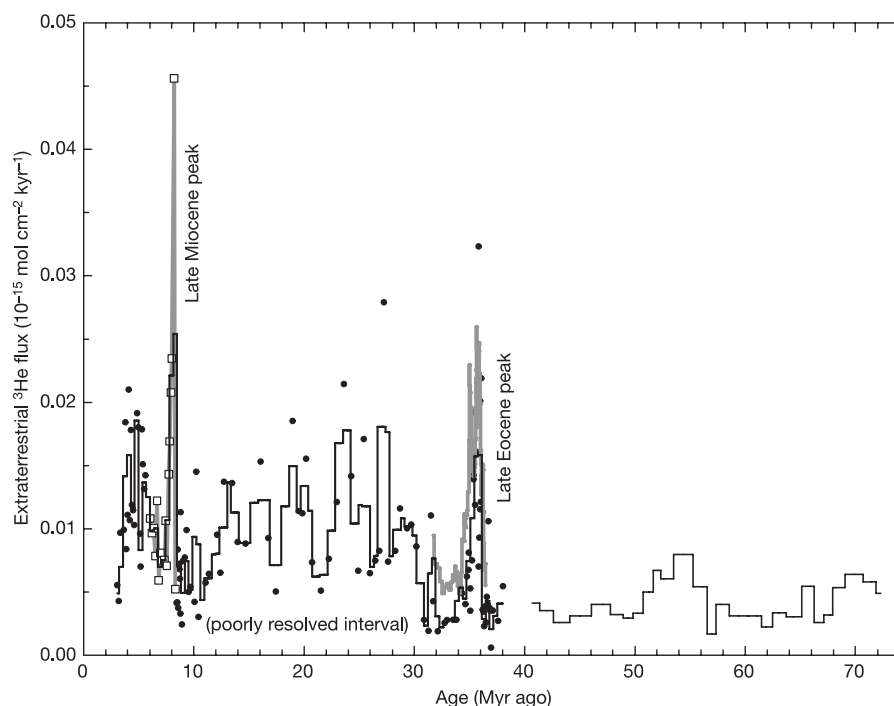


Figure 1 | Two periods of elevated ^3He flux, at ~ 35 Myr ago (late Eocene) and ~ 8 Myr ago (late Miocene), indicate intervals of enhanced accumulation rate of IDPs. Open and filled symbols are individual new ^3He measurements from ODP Site 757 (central Indian Ocean, $17^\circ 01.458' \text{ S}$, $88^\circ 10.899' \text{ E}$). Lines are 3-point running means through the data points, taken to minimize the effects of occasional sampling of large individual

IDPs¹⁰. The grey segments of the running mean line indicate the late Miocene event (highlighted by open symbols), and the previously reported late Eocene peak from the Italian Apennines¹. Cretaceous to mid-Tertiary running mean data are from ref. 10. Details of the new samples, analytical methods, data and age models are provided in Supplementary Information.

asteroid shower are ~ 0.1 AU from the Veritas family¹⁶ (for example, the 9:4, 11:5 or 2:1 mean motion resonances with Jupiter). This distance would either require huge ejection velocities from Veritas (that are not observed) or extremely long drift times via Yarkovsky thermal forces (which would fail to produce a spike of impactors). (Yarkovsky drift occurs when small asteroids absorb solar photons, heat up and then reradiate the energy away in a non-isotropic manner after a short delay⁷.) Moreover, these resonances are very unlikely to produce Earth impactors^{17,18}. Thus, the event that formed the Veritas family almost certainly did not produce an asteroid shower on Earth, so the absence of craters of this age is not surprising.

To investigate under what conditions the Veritas collision might produce a dust spike similar to that defined by the ^3He record, we developed a statistical Monte Carlo model to track the collisional and dynamical evolution of particles formed by the disruption of Veritas. (Model details can be found in Supplementary Information). The results of our code were calibrated by modelling the evolution of dust in several latitudinal bands observed by the Infrared Astronomical Satellite (IRAS; see ref. 15 for details). To compare our model results with the ^3He data, we calculated the flux of $D = 10 \mu\text{m}$ particles reaching 1 AU. Particles of about this size can escape intense atmospheric entry heating and He loss, and currently dominate the ^3He flux to Earth⁸. Since we do not yet have a model of ^3He implantation in IDPs, or one for heating and helium retention during atmospheric entry of IDPs produced during the dust spike, here we simply compare our model 1 AU flux with the shape and duration of the ^3He peak.

Small particles from the Veritas family were assumed to reach Earth through P-R and solar wind drags. The production rate of the first-generation particles that were started at 3.17 AU was defined using a 'broken' cumulative power-law size frequency distribution (SFD) with two slopes, index α_1 at smaller sizes and α_2 at larger sizes. The SFD extends from $D = 10 \mu\text{m}$ to 1 cm. We assume the SFD decays exponentially with time from collisional evolution and radiation drag forces. This causes the diameter of the knee between

α_1 and α_2 to increase with time. The rate of IDP disruptions was defined as a function of diameter D and heliocentric distance R (ref. 19). When a particle disrupts, we replace it with a swarm of fragments that follow a power-law SFD. We assumed that the mass of the largest fragment was half that of the parent particle. The power law index of the fragment size distribution was determined by mass conservation²⁰. Thus, in our simulations, we follow several generations of particles produced by a collisional cascade; typical runs track the histories of 10^8 – 10^9 particles.

We find that $D = 10 \mu\text{m}$ IDPs from Veritas reach 1 AU in ~ 40 kyr, typically shorter than their collisional lifetime (20–200 kyr, depending on model assumptions¹⁹). This means the ^3He signal at Earth would be extremely brief unless these particles are continuously replenished in some fashion. The Veritas break-up, however, produced a SFD of fragments. We note that $D = 1$ – 5 mm particles have short collisional lifetimes¹⁹ (< 100 kyr), such that their fragments not only replenish the $D = 10 \mu\text{m}$ population but also create intermediate-size particles that also dynamically evolve and disrupt over time. This means that the ^3He signal was produced by fragments from a collisional cascade that was fed new material by disruption events occurring both near the Veritas source region and *en route* to Earth.

The late Miocene event allows us to glean insights into the SFD of particles produced by the break-up and how it changes with time. We found that the initial break-up of the Veritas family probably produced a swarm of IDPs that dominated the main belt population by at least an order of magnitude for ~ 1 Myr. Our best fit to the shape and the decay time of the ^3He peak comes from using $\alpha_1 = -2.5$ and $\alpha_2 = -3.3$ (Fig. 2). We found that α_2 values significantly shallower or steeper than our best fit value produce ^3He peaks that are longer or shorter, respectively, than those observed.

Today, collisions in the Veritas family produce one of the prominent dust bands observed by infrared telescopes, and also contribute at least 5×10^6 kg per year to the terrestrial IDP flux¹⁵. Our modelling suggests that the IDP flux from the Veritas family will continue to decay for several tens of Myr until it reaches a collisional steady state

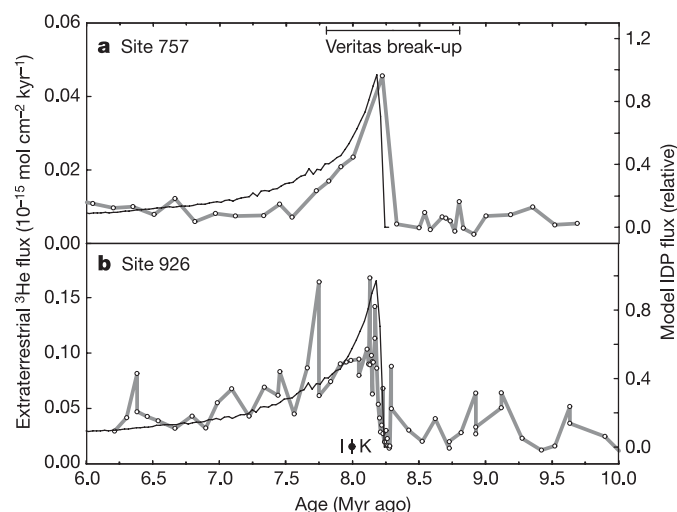


Figure 2 | Concordant ^3He data from two widely separated seafloor sites support a global increase in IDP flux at 8.2 Myr ago that can be attributed to the asteroid collision that produced the Veritas family. **a, b,** The extraterrestrial ^3He flux peak (grey line, with small symbols) through the Late Miocene event is similar at Site 757 (**a**) and Site 926 (**b**; western equatorial Atlantic; $3^\circ 43.148' \text{ N}$, $42^\circ 54.507' \text{ W}$). The modelled $10\text{-}\mu\text{m}$ IDP flux following the Veritas collision is shown by the black curves. The model dust spike was positioned at 8.25 Myr ago and scaled to align with the ^3He peak. The fast rise time and ~ 1.5 Myr decay time observed in the ^3He record at both sites are well matched by the model. The inferred time of Veritas break-up is indicated (see Supplementary Fig. S4 for details). The symbol at ~ 8 Myr ago indicates transition from illite (I) to kaolinite (K) clays at Site 926, a proxy for climate change.

and takes on the same approximate shape of the overall main belt size distribution for $D < 5$ km bodies²¹.

The late Miocene event roughly coincides with cosmic ray exposure ages of 7–8 Myr on many H-chondrite meteorites²². The connection between these observations, however, is unclear. As described previously, the nearest powerful resonances are not only ~ 0.1 AU from Veritas family members, but they are also highly inefficient at producing Earth impactors. Moreover, mineralogical and spectroscopic differences between Veritas family members and the H-chondrites indicate that the latter almost certainly did not originate on the former²³. We also find it highly unlikely that the projectile that produced the Veritas family was the source of the H-chondrites, partly for the reasons above, but also because the H-chondrites do not show evidence for significant shocks at 7–8 Myr ago²⁴. If the two events are indeed linked, we postulate that Veritas family members disrupted a well-positioned fragment from the H-chondrite parent body shortly after the family-forming event took place.

Previous work has suggested a possible link between the IDP accretion rate and Earth's climate²⁵. Correlations between extraterrestrial ^3He in sediments and global climate in the Quaternary period may support this suggestion⁹ but also may be an artefact of climate-induced changes in sedimentation²⁶. Modest global cooling and strengthening of the Asian monsoon occurred in the late Miocene²⁷. At Site 926 there is a sharp transition from kaolinite-rich to illite-rich sediment²⁸, occurring within the ^3He peak but 200 kyr after its onset (Fig. 2). This transition may document a change from warm humid to cold dry continental weathering. Although the relative timing of these events is suggestive, we caution that a compelling link between the events cannot be established until a plausible mechanism is found by which IDPs can change climate.

Received 21 June; accepted 31 October 2005.

1. Farley, K. A., Montanari, A., Shoemaker, E. M. & Shoemaker, C. S. Geochemical

evidence for a comet shower in the Late Eocene. *Science* **280**, 1250–1253 (1998).

2. Takayanagi, M. & Ozima, M. Temporal variation of $^3\text{He}/^4\text{He}$ in deep-sea sediment cores. *J. Geophys. Res.* **92**, 12531–12538 (1987).
3. Farley, K. A. Cenozoic variations in the flux of interplanetary dust recorded by ^3He in a deep-sea sediment. *Nature* **376**, 153–156 (1995).
4. Kortenamp, S. & Dermott, S. A 100,000-year periodicity in the accretion rate of interplanetary dust. *Science* **280**, 874–876 (1998).
5. Nesvorný, D., Bottke, W. F., Levison, H. F. & Dones, L. Recent origin of the solar system dust bands. *Astrophys. J.* **591**, 486–497 (2003).
6. Tagle, R. & Claeys, P. Comet or asteroid shower in the late Eocene? *Science* **305**, 492 (2004).
7. Burns, J. A., Lamy, P. L. & Soter, S. Radiation forces on small particles in the Solar-System. *Icarus* **40**, 1–48 (1979).
8. Farley, K. A., Love, S. G. & Patterson, D. B. Atmospheric entry heating and helium retentivity of interplanetary dust particles. *Geochim. Cosmochim. Acta* **61**, 2309–2316 (1997).
9. Patterson, D. B. & Farley, K. A. Extraterrestrial He-3 in seafloor sediments: Evidence for correlated 100 kyr periodicity in the accretion rate of interplanetary dust, orbital parameters, and Quaternary climate. *Geochim. Cosmochim. Acta* **62**, 3669–3682 (1998).
10. Mukhopadhyay, S., Farley, K. & Montanari, A. A 35 Myr record of helium in pelagic limestones: implications for interplanetary dust accretion from the early Maastrichtian to the Middle Eocene. *Geochim. Cosmochim. Acta* **65**, 653–669 (2001).
11. Shackleton, N. J., Curry, W. B., Richter, C. & Bralower, T. J. Ceara Rise. *Proc. ODP Sci. Res.* **154**, 1–552 (1997).
12. Bottke, W. F. et al. The fossilized size distribution of the main asteroid belt. *Icarus* **175**, 111–140 (2005).
13. Low, F. J. et al. Infrared cirrus—new components of the extended infrared-emission. *Astrophys. J.* **278**, L19–L22 (1984).
14. Dermott, S. F., et al. in *Interplanetary Dust* (eds Grün, E., Gustafson, B. A. S., Dermott, S. F. & Fechtig, H.) 569–639 (Springer, Berlin, 2001).
15. Nesvorný, D., Vokrouhlický, D., Bottke, W. F. & Sykes, M. Physical properties of asteroid dust bands and their sources. *Icarus* (submitted).
16. Morbidelli, A. & Nesvorný, D. Numerous weak resonances drive asteroids toward terrestrial planet orbits. *Icarus* **139**, 295–308 (1999).
17. Gladman, B. J. et al. Dynamical lifetimes of objects injected into asteroid belt resonances. *Science* **277**, 197–201 (1997).
18. Bottke, W. F. et al. Debaised orbital and absolute magnitude distribution of the near-earth objects. *Icarus* **156**, 399–433 (2002).
19. Grün, E., Zook, H. A., Fechtig, H. & Giese, R. H. Collisional balance of the meteoritic complex. *Icarus* **62**, 244–272 (1985).
20. Nolan, M. C. & Greenberg, R. Stochastic-evolution of asteroids to produce the ordinary chondrites. *Meteoritics* **24**, 310 (1989).
21. Bottke, W. F. et al. in *Dynamics of Populations of Planetary Systems* (eds Knezevic, Z. & Milani, A.) 357–376 (IAU Colloquium 197, Cambridge Univ. Press, Cambridge, UK, 2005).
22. Graf, T. & Marti, K. Collisional history of H chondrites. *J. Geophys. Res.* **100**, 21247–21263 (1995).
23. Burbine, T. H., McCoy, T. J., Meibom, A., Gladman, B. & Keil, K. in *Asteroids III* (eds Bottke, W. F., Cellino, A., Paolicchi, P. & Binzel, R. P.) 653–667 (Univ. Arizona Press, Tucson, 2002).
24. Bogard, D. D. Impact ages of meteorites—A synthesis. *Meteoritics* **30**, 244–268 (1995).
25. Muller, R. A. & MacDonald, G. J. Glacial cycles and astronomical forcing. *Science* **277**, 215–218 (1997).
26. Marcantonio, F. et al. Extraterrestrial ^3He as a tracer of marine sediment transport and accumulation. *Nature* **383**, 705–707 (1996).
27. Gupta, A. K., Singh, R. K., Joseph, S. & Thomas, E. Indian Ocean high-productivity event (10–8 Ma): Linked to global cooling or to the initiation of the Indian monsoons? *Geology* **32**, 753–756 (2004).
28. Curry, W. B., et al. *ODP Init. Rep. Leg 154* 1–1111, (1995).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements Financial support for this project was provided by NASA's Planetary Geology & Geophysics program (W.F.B., D.N. and K.A.F.). Financial and travel support for D.V. was provided by the Czech Republic grant agency and NSF's COBASE program. We also thank D. Durda, A. Morbidelli, M. Sykes and S. Mukhopadhyay for several discussions, and S. Goldstein and J. Burns for comments and suggestions.

Author Contributions K.A.F. measured ^3He in the seafloor sediments. D.N. determined the age of the Veritas family using numerical integration methods. D.V., W.F.B. and D.N. constructed the Monte Carlo dust evolution code and analysed the results.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to K.A.F. (farley@gps.caltech.edu).

LETTERS

Proton–proton correlations observed in two-proton radioactivity of ^{94}Ag

Ivan Mukha^{1,2,3,†}, Ernst Roeckl¹, Leonid Batist⁴, Andrey Blazhev^{1,5}, Joachim Döring¹, Hubert Grawe¹, Leonid Grigorenko⁶, Mark Huyse², Zenon Janas⁷, Reinhard Kirchner¹, Marco La Commara⁸, Chiara Mazzocchi¹, Sam L. Tabor⁹ & Piet Van Duppen²

The stability and spontaneous decay of naturally occurring atomic nuclei have been much studied ever since Becquerel discovered natural radioactivity in 1896. In 1960, proton-rich nuclei with an odd or an even atomic number Z were predicted¹ to decay through one- and two-proton radioactivity, respectively. The experimental observation of one-proton radioactivity was first reported² in 1982, and two-proton radioactivity has now also been detected by experimentally studying the decay properties of ^{45}Fe (refs 3, 4) and ^{54}Zn (ref. 5). Here we report proton–proton correlations observed during the radioactive decay of a spinning long-lived state of the lightest known isotope of silver⁶, ^{94}Ag , which is known to undergo one-proton decay⁷. We infer from these correlations that the long-lived state must also decay through simultaneous two-proton emission, making ^{94}Ag the first nucleus to exhibit one- as well as two-proton radioactivity. We attribute the two-proton emission behaviour and the unexpectedly large probability for this decay mechanism to a very large deformation of the parent nucleus into a prolate (cigar-like) shape, which facilitates emission of protons either from the same or from opposite ends of the ‘cigar’.

The recent observation^{3,4} of two-proton radioactivity (2p) for ^{45}Fe yielded a half-life of about 4 ms, which exceeds by a factor of a thousand the quasi-classical estimates for ‘di-proton’ (or ^2He -cluster) emission from the ^{45}Fe nucleus. But the value is in accord with quantum-mechanical theory^{8,9} that assumes a three-body structure (two protons and a remnant core) for the parent nucleus, with resultant high Coulomb and centrifugal barriers slowing down this exotic decay. Long-lived 2p precursors may thus be fairly common^{10,11}, including 2p radioactivity from isomeric (that is, metastable) nuclear states. However, we will consider here only the radioactivity of nuclei that live long enough to establish the atomic structure of the corresponding chemical elements¹². That is, we refrain from discussing 2p emission from short-lived nuclear resonances^{13–18} with half-lives that are so short that the mechanism of the 2p-emission depends on the way in which the parent resonance is populated, in contrast to 2p radioactivity.

Until now, experimental data on 2p decays^{3–5} have been interpreted as being dominated by single-proton configurations. In contrast, the present work uses a nucleus with radioactive decay characteristics that call for a consideration of collective structures. More specifically, we study the metastable state (isomer) of ^{94}Ag with the tentative spin (nuclear angular momentum) and parity assignment $(21^+)^{19,20}$, which is a nuclear ‘spin trap’ with the largest spin ever observed for β -decaying nuclei. It belongs to a sequence of states with a collective structure in which valence protons are coupled with

neutrons into deuteron-like pairs²⁰. The spin of such levels can vary within a broad range and usually increases with excitation energy as more nucleons have to be aligned in spin, which requires breaking of anti-aligned pairs of nucleons. However, in ^{94}Ag an inversion in the normal ordering of the 19^+ and 21^+ levels occurs, owing to additional correlation energy, so that the most probable γ -ray de-excitation—from the (21^+) state to the (17^+) neighbour level—is very slow. Other decay modes therefore take over, which results in the long half-life value²¹ of 0.39(4) s. This isomer has unique properties unmatched in the entire nuclear chart: although its dominant radioactivity mode is β -decay followed by γ -ray or proton emission^{20,21}, the excitation energy⁷ of 6.7(5) MeV makes three decay modes possible: one-proton, α -particle and two-proton radioactivity²².

The evidence for 2p decay of the (21^+) isomer of ^{94}Ag to an excited state of the daughter nucleus ^{92}Rh was obtained by analysing the same experimental data used to document its one-proton radioactivity⁷. Whereas earlier 2p-decay measurements^{3–5} were based on single decay-energy spectra and related to the decay of the daughter nucleus, we searched for a 2p-radioactivity branch by looking for two individual protons and two γ -rays, all occurring in coincidence with each other. Tagging these events by the known γ transitions²³ in ^{92}Rh then provides an unambiguous signature of the 2p decay of the (21^+) isomer in ^{94}Ag (see Fig. 1 and also Methods and Supplementary Information). The 2p branch is characterized by a decay energy of 1.9(1) MeV and a decay probability of 0.5(3)% leading to a partial half-life of $80(^{+110}_{-30})$ s. The derived tentative decay scheme is shown in Fig. 2. The angular momentum, L , involved in the emission of the two protons is estimated to be between 6 and 10 units (for details, see Methods and Supplementary Information).

In general, 2p decay can proceed through sequential proton emission involving intermediate ^{93}Pd states, or through a simultaneous three-particle decay mechanism. The present experiment is accordingly designed to observe two coincident protons and their angular and energy correlations. Although the recoiling ^{92}Rh partner in the three-body break-up remains undetected, these data allow us to extract information about the decay mechanism¹¹. The angular and energy correlations are apparent in relative energy spectra, such as those showing the relative energy of the two emitted protons ($E_{\text{p-p}}$ in Fig. 3a) and the energy of one emitted proton relative to that of the ^{92}Rh daughter ($E_{\text{p-Rh}}$ in Fig. 3b). The relative-energy spectra were calculated according to standard kinematics formulae²⁴ (see also Supplementary Information); for every pair of silicon detectors (fifteen different detector combinations in total) we used the average emission angle and corresponding detection probability indicated by

¹Gesellschaft für Schwerionenforschung, D-64291 Darmstadt, Germany. ²Instituut voor Kern- en Stralingsfysica, KU Leuven, B-3001 Leuven, Belgium. ³Kurchatov Institute, RU-123184 Moscow, Russia. ⁴St Petersburg Nuclear Physics Institute, RU-188350 Gatchina, Russia. ⁵University of Sofia, BG-1164 Sofia, Bulgaria. ⁶Flerov Laboratory of Nuclear Reactions, JINR, RU-141980 Dubna, Russia. ⁷Warsaw University, PL-00681 Warsaw, Poland. ⁸Università “Federico II” and INFN Napoli, I-80126 Napoli, Italy. ⁹Florida State University, FL-32306 Tallahassee, USA. [†]Present address: University of Seville, ES-41080 Seville, Spain.

Monte Carlo simulations that assume an isotropic emission probability for the two protons.

The relative energy spectrum E_{p-p} in Fig. 3a exhibits two distinct peaks, separated by a region at intermediate energies where no counts are recorded. The low-energy E_{p-p} peak corresponds to protons emitted in almost the same direction, whereas the high-energy peak corresponds to protons emitted in the opposite direction. This E_{p-p} distribution rules out the sequential emission mechanism that should result in a uniform distribution

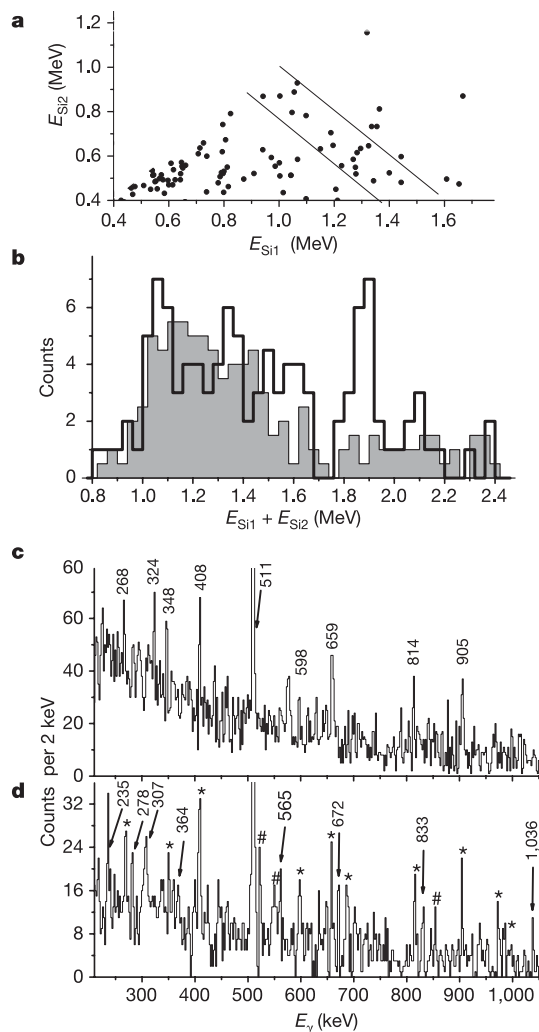


Figure 1 | Evidence for two-proton decay of $^{94}\text{Ag}(21^+)$. **a**, Observed energy correlations of protons. The spectrum is obtained in a $\text{Si}_1\text{-Si}_2\text{-}\gamma\text{-}\gamma$ coincidence matrix with double γ -ray gates on ^{92}Rh γ -transitions. Experimental energy thresholds are ~ 400 keV. **b**, Sum-energy spectra of two Si detectors. They are obtained in $\text{Si}_1\text{-Si}_2\text{-}\gamma\text{-}\gamma$ coincidence with double γ -ray gates on ^{92}Rh (bold histogram), and with double γ -gates-shifted left and right off the respective ^{92}Rh transitions by 3 keV (filled histogram). The only significant difference between the two histograms is the peak at 1.9 MeV, interpreted as the 2p decay $^{94}\text{Ag}(21^+) \rightarrow ^{92}\text{Rh}^* + p + p$. The corresponding region in the $\text{Si}_1\text{-Si}_2$ scatter plot (**a**) is given by two parallel lines. The broad bump around 1.2 MeV is due to re-scattered β particles that hit two Si detectors and are coincident with intense Compton-scattered γ rays from the β decay of ^{94}Ag , thus imitating the proton–proton– $\gamma\text{-}\gamma$ events. **c**, **d**, The observed γ -ray spectra. They are projected out of the $\text{Si}_1\text{-Si}_2\text{-}\gamma\text{-}\gamma$ coincidence matrix with a double condition: the energy of the eight γ -ray transitions in ^{92}Rh as first condition on one γ -ray detector, and with the $\text{Si}_1\text{-Si}_2$ sum-energy applied as the second condition in the ranges 1.2–1.6 and 1.8–1.95 MeV, respectively. Known transitions in ^{94}Pd , ^{93}Rh and ^{92}Rh are marked by symbols (*, # and arrow, respectively), and/or with their energies in keV.

of E_{p-p} values (see ref. 25), with a probability of more than 95% (for further discussion, see Supplementary Information). Moreover, the E_{p-Rh} spectrum displayed in Fig. 3b does not exhibit narrow peaks owing to the population of intermediate ^{93}Pd states, and there is no evidence for energy correlations between the two emitted protons (see $E_{\text{Si}1}\text{-}E_{\text{Si}2}$ plot in Fig. 1a), as would be expected if sequential proton emission were taking place (see also discussion in Supplementary Information). Sequential emission might be considered feasible when bearing in mind that the recently observed one-proton radioactivity of the (21^+) isomer⁷ populates two excited states of ^{93}Pd with probabilities of 1.9(5)% and 2.2(4)%, with these excited states in principle able to undergo further one-proton decay; overall, such a process would appear in our data much like a 2p process. However, the two excited ^{93}Pd states both de-excite by fast γ -decay, leaving no chance for further proton emission⁷.

Taken together, our observations seem consistent only with ^{94}Ag decaying through a simultaneous 2p emission process. For ^{94}Ag such a ‘true’ 2p-decay mode should be much slower than the sequential decay mode⁹; moreover, the 2p-decay half-life estimated with the ‘simultaneous emission’ model^{10,11} exceeds the experimentally determined half-life by a factor of $10^3\text{--}10^6$. The unexpectedly large 2p decay probability we observe can, however, be explained by assuming strong nuclear deformation of the (21^+) isomer. We used a simple model (see Methods) to evaluate the probability of simultaneous 2p decay of a deformed nucleus, assuming that the two protons share the decay energy E and angular momentum L and that they are emitted from different points on an elliptically deformed nucleus with axes ratios of 1:1: a . The results of these calculations (which do not include proton–proton interactions) are shown in Fig. 4a for $L = 6, 8$ and 10, and are seen to reproduce the experimental data when using nuclear deformation parameters $a < 0.3$ or $a > 2$ (corresponding to oblate or disk-like deformation, and prolate or cigar-like deformation, respectively). The extent of the nuclear deformation that is needed to replicate the experimental half-life is very large. For example, oblate nuclei with $a \approx 0.3$ have not been observed so far, whereas prolate deformations with $a \approx 2$ have only been seen in super-deformed nuclei²⁶. For comparison, Fig. 4a also includes the predictions of the quasi-classical ^2He model¹ that assumes a strong proton–proton attraction resulting in the emission of a ^2He cluster from the deformed nuclear surface. In the case of spherical (that is, not deformed) nuclei, the ^2He model provides a lower limit for 2p-decay half-lives^{10,11} and is used here for reference purposes.

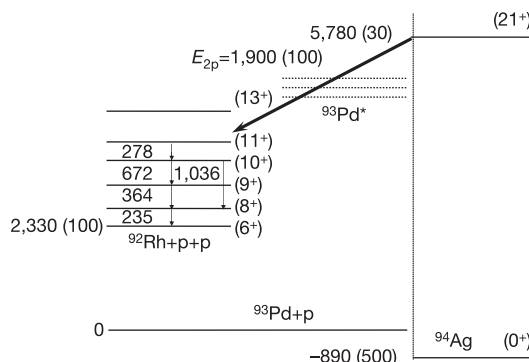


Figure 2 | Two-proton decay of the $^{94}\text{Ag}(21^+)$ isomer to an excited state in ^{92}Rh . All levels are given with respect to the ^{93}Pd ground state. The ^{92}Rh level energies²³, γ -ray transitions of ^{92}Rh and the excitation energy of the (21^+) state⁷ are in keV. The triple dotted lines indicate excited states of ^{93}Pd that might be involved in sequential 2p decay. On the far left side, the proton separation energy in ^{93}Pd is indicated as tentatively assigned in this work. The energy of the ^{94}Ag ground state stems from the recent atomic mass evaluation²². Note that all spin assignments are tentative and are based on previous work^{20–23}.

Using a prolate deformation of $a \approx 2$, our model also qualitatively describes the observed proton–proton correlations in the E_{p-p} spectrum of Fig. 3a. The protons are emitted with the largest probability within narrow cones around the poles of the ellipsoid (see Fig. 4b), either from the same pole or from opposite ones. The most spectacular influence of the deformed nuclear shape on the proton–proton correlations is the back-to-back emission, where the protons are ejected from opposite sides of the parent nucleus yet still exhibit strongly correlated energies. The quasi-classic ${}^2\text{He}$ model cannot explain the experimentally observed proton–proton correlations that suggest back-to-back proton emission. Comparing the experimental data with the calculation, 2p decay from the same pole (left peak in Fig. 3a) seems to be enhanced, which supports predictions that the relative-energy proton–proton spectrum should be enhanced at low energies due to proton–proton attraction²⁷. The relative-energy spectrum for proton– ${}^{92}\text{Rh}$ correlations appears symmetric around the half-decay-energy value where the energies of both protons are equal (see Fig. 3b). The simultaneous-emission model can reproduce this feature qualitatively, but the predictions are not sensitive to the deformation parameter. The almost-equal energies of

both protons were predicted to be experimental signature of the ‘true’ 2p decay of spherical nuclei^{1,10,11}.

Two-proton decay was predicted to be observable in proton-dripline nuclei with an even number of protons (Z) only, because the strength of the proton–proton attraction makes them more bound than their odd- Z neighbours. The observation of 2p radioactivity for the (21^+) isomer of ${}^{94}\text{Ag}$ is thus surprising, considering that it is an odd- Z nucleus with a structure presumably dominated by deuteron-like proton–neutron pairs. The present observations are also unusual in that they make the (21^+) isomer the first long-lived nuclear state in which both one- and two-proton radioactivity has been identified. The data can be explained qualitatively with a simple model that assumes very strong nuclear deformation as the collective phenomenon that determines the 2p-decay pattern. But whereas this model can reproduce proton–proton correlations and explain the high probability for proton emission from the compact polar regions of the nucleus, developing a quantitative description of the observed 2p decay behaviour remains a challenge to theory. Attempts to further our understanding of the decay behaviour of the (21^+) isomer of ${}^{94}\text{Ag}$ will no doubt benefit from experiments that aim to obtain proton–proton correlations with high resolution and better statistics and directly to determine the shape of the (21^+) isomer via measurements of the nuclear quadrupole moment.

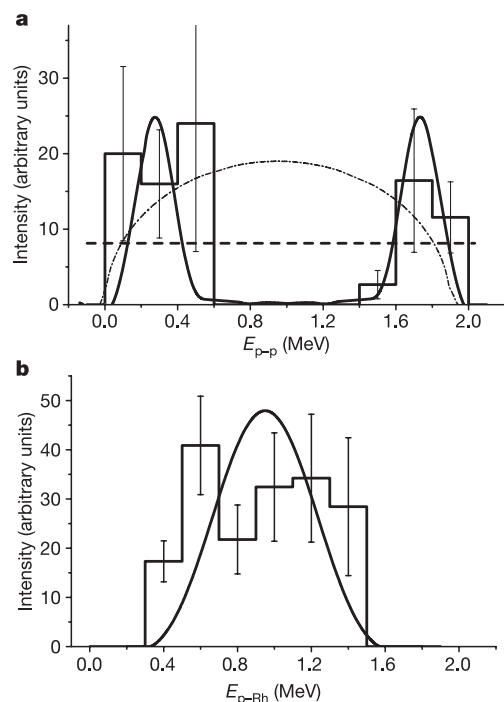


Figure 3 | Correlations observed in the 2p decay of ${}^{94}\text{Ag}(21^+)$. Relative-energy spectra for proton–proton (E_{p-p}) and proton– ${}^{92}\text{Rh}$ (E_{p-Rh}) correlations are shown by the histograms in **a** and **b**, respectively. The spectra were derived out of the $\text{Si}_1\text{--Si}_2\text{--}\gamma\text{--}\gamma$ coincidence matrix, with a total of 19 events fulfilling the triple condition set on two ${}^{92}\text{Rh}$ γ -transitions and on the 2p sum-energy in the range of 1.8–1.95 MeV. All proton–proton events are normalized by using their respective solid angles of the detectors. This leads to a normalization of the histogram and therefore the intensity is given in arbitrary units. The experimental uncertainties (shown as error bars) are calculated as weighted sums of the standard deviations of the respective proton–proton counts. The spectra were derived from the raw data displayed in Supplementary Fig. S2. The solid curves are the predictions of our model of simultaneous proton emission from a deformed nucleus convoluted with an experimental uncertainty of 200 keV. The dashed line represents the fit obtained with a sequential emission mechanism. The latter decay should result in two narrow peaks in **b** whose energies depend on the (unknown) energy of the ${}^{93}\text{Pd}$ state involved and should add up to a total of 1.9(1) MeV. As there is no evidence for such pair(s) of peaks in **b**, their positions are not indicated. The dashed-dotted curve shows the calculated distribution when the 2p decay is isotropic in the absence of the mentioned decay mechanisms.

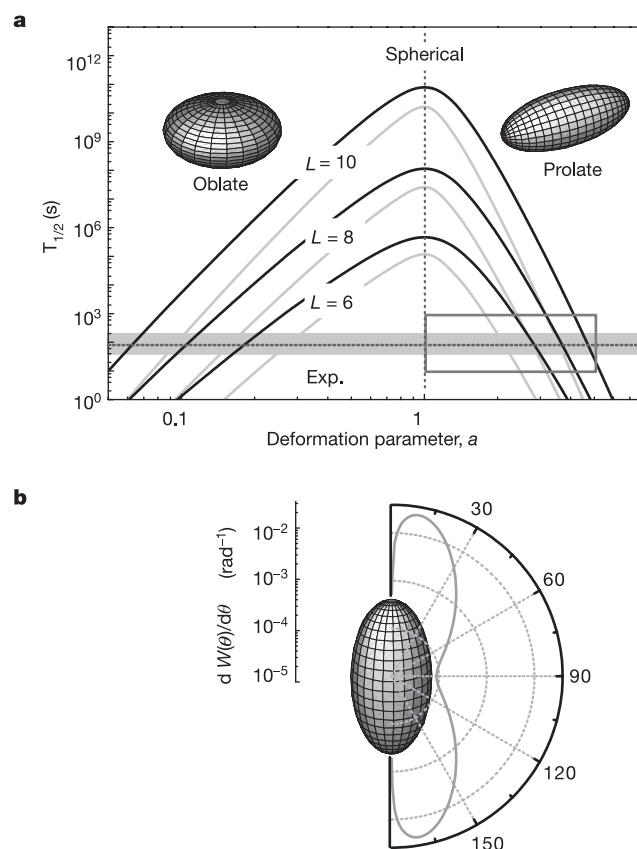


Figure 4 | Nuclear deformation of ${}^{94}\text{Ag}(21^+)$ derived from its two-proton decay. **a**, Partial 2p-decay half-life ($T_{1/2}$) of ${}^{94}\text{Ag}(21^+)$ as a function of the nuclear deformation parameter a , which is a ratio of the long to the short axes of the ellipsoid. The simultaneous (black curves) and quasi-classical ${}^2\text{He}$ (grey curves) model estimates of the 2p decay are shown for the angular momenta $L = 6, 8, 10$. The nuclear shapes corresponding to the derived deformations are shown for the $L = 6$ calculation. The experimental $T_{1/2}$ value is shown by the dotted line (marked as ‘Exp.’) with the grey region of experimental errors. **b**, Intensity W of 2p-emission as a function of the angle θ between one emitted proton and the long ellipsoid axis (thick grey curve).

METHODS

Identification of two-proton radioactivity. The measurements were performed at the on-line mass separator²⁸ in Gesellschaft für Schwerionenforschung (GSI), Darmstadt, Germany. By using the ⁵⁸Ni(⁴⁰Ca, p3n) fusion–evaporation reaction and the isotope-separation on-line (ISOL) ion source²⁹, the ⁹⁴Ag isomers were produced with an intensity of 2 atoms s^{−1}. The isomer decay products were measured by an array of six charged-particle Si and seventeen γ-ray Ge detectors providing complete kinematics information about the decays. Further details of the experiment are given in ref. 21 and in the Supplementary Information.

We searched for 2p decay by measuring fourfold coincidences between double-hit charged-particle events recorded by the Si detectors and γ–γ coincidence events registered by the Ge detectors. The γ-ray gates were set on known ⁹²Rh transitions²³. The derived sum–energy spectra ($E_{\text{Si1}} + E_{\text{Si2}}$) are displayed in Fig. 1b. The histogram shows a statistically significant peak (19 events), indicating 2p decay with an $E_{\text{Si1}} + E_{\text{Si2}}$ energy of 1.88(9) MeV and a branching ratio of 0.5(3)%. The latter quantity was determined by comparing the proton–proton–γ–γ coincidence intensity with the β–delayed proton branches of the ⁹⁴Ag isomers²¹. The corresponding 2p-decay energy is 1.9(1) MeV. To confirm these conclusions, we inspected the γ-ray data conditioned with energy gates in the $E_{\text{Si1}} + E_{\text{Si2}}$ spectrum both ‘on’ and ‘off’ the 1.9 MeV peak. The spectrum of γ-rays following the 1.9 MeV 2p decay (see Fig. 1d) shows the six known γ-transitions in ⁹²Rh. In addition, we observed γ-rays at 565 and 833 keV, which are also coincident with known γ transitions in ⁹²Rh but have not been identified by in-beam γ-spectroscopy²³. Our data yield evidence that the 1.9 MeV peak coincides with all γ-rays following the de-excitation of the known (11⁺) state in ⁹²Rh, except for the 307 keV γ-transition, whose position in the ⁹²Rh level scheme is uncertain²³. No evidence for population of the higher-lying (13⁺) or (11[−]) states in ⁹²Rh was obtained, because the corresponding γ transitions were not observed; see the spectrum in Fig. 1d. Therefore we very tentatively assign the observed 2p decay to feed the (11⁺) level in ⁹²Rh, although we cannot exclude 2p decay to an unknown state in ⁹²Rh, which feeds the (11⁺) level and might also de-excite by the unplaced 307, 565 or 833 keV γ-transitions. Thus the orbital angular momentum of 2p decay is conservatively estimated to be between 6 and 10 units by assuming that un-assigned γ-rays might take up to 4 units of angular momentum.

The γ-spectrum from the fourfold coincidence events, obtained by gating on $E_{\text{Si1}} + E_{\text{Si2}}$ energies in the region below the 1.9 MeV peak, is shown in Fig. 1c. It is dominated by the known ¹⁹²⁰ γ-transitions in ⁹⁴Pd following the β decay of ⁹⁴Ag. This observation explains the broad bump of the sum-energy spectrum shown in Fig. 1b. In a similar way, a γ-ray spectrum was generated with the $E_{\text{Si1}} + E_{\text{Si2}}$ condition shifted to energies above the 1.9 MeV peak. This spectrum (not shown) is dominated by the known ⁹³Rh γ-transitions populated by β-delayed proton emission²¹ from ⁹⁴Ag. Some intense ⁹⁴Pd and ⁹³Rh γ transitions are seen in the spectrum in Fig. 1d. They can be explained by the background under the 1.9 MeV peak, owing to re-scattered β particles. However, the ⁹²Rh γ-transitions are absent in both γ-spectra generated by gating below and above the 2p peak (see Fig. 1c). A final check for possible ‘imitation’ of 2p events was performed by generating a $E_{\text{Si1}} + E_{\text{Si2}}$ spectrum from the measured Si₁–Si₂–γ–γ coincidence with the γ-energy gates being shifted 3 keV above and under the nominal value of ⁹²Rh transitions. The resulting $E_{\text{Si1}} + E_{\text{Si2}}$ spectrum displayed in Fig. 1b does not show any indication of the 1.9 MeV peak providing the background estimate. More details are given in Supplementary Information.

Model of two-proton emission from a deformed nucleus. Two protons are assumed to share the decay energy E and angular momentum L , and to be emitted independently from different points on an axially deformed nuclear surface in the radial direction. This surface is given by a radius $r(\theta) = (a^2 A)^{1/3} r_0 / \sqrt{1 + (a^2 - 1) \sin^2(\theta)}$, with A being the number of nucleons and $r_0 = 1.45$ fm (1 fm = 10^{−15} m) being the reduced radius. The volume of such an ellipsoid, $4\pi A r_0^3/3$, does not depend on the deformation parameter a , which is equal to the long-to-short axis ratio. The 2p-decay width can then be estimated as:

$$\begin{aligned} \Gamma &= \frac{\hbar \cdot \ln 2}{T_{1/2}} = \int_0^\pi \frac{d\Gamma(\theta)}{d\theta} d\theta \\ &= \Gamma_w \sum_{\ell=0}^L (1 + \delta'_{L-\ell}) \int_0^E d\varepsilon \int_0^1 d\cos(\theta) \int_0^1 d\cos(\theta') \\ &\quad P_\ell(\varepsilon, r(\theta), Z - 2) P_{L-\ell}(E - \varepsilon, r(\theta'), Z - 2) \end{aligned}$$

where P_ℓ is the standard penetrability, which depends on the energy ε and orbital

angular momentum l of the proton³⁰, and $\delta'_l = 1$ or otherwise 0. The channel radius is assumed to be $r(\theta)$. The upper (Wigner) limit of the 2p-decay width is calculated as for spherical nuclei^{10,11}, $\Gamma_w = 6S_{2p}(\pi E^{1/2} M^{3/2} r_0^3 A)^{-1}$. The preformation factor S_{2p} , related to the spectroscopic factor of the 2p configuration in the parent nucleus, was assumed to be unity, and M stands for the mass of the nucleon. The differential probability of a proton emission is defined as $\frac{dW(\theta)}{d\theta} = \frac{1}{T} \frac{d\Gamma(\theta)}{d\theta}$.

Received 26 May; accepted 21 November 2005.

- Goldansky, V. I. On neutron-deficient isotopes of light nuclei and the phenomena of proton and two-proton radioactivity. *Nucl. Phys.* **19**, 482–495 (1960).
- Hofmann, S. *et al.* Proton radioactivity of ¹⁵¹Lu. *Z. Phys. A* **305**, 111–115 (1982).
- Pfützner, M. *et al.* First evidence for the two-proton decay of ⁴⁵Fe. *Eur. Phys. J. A* **14**, 279–285 (2002).
- Giovannazzo, J. *et al.* Two-proton radioactivity of ⁴⁵Fe. *Phys. Rev. Lett.* **89**, 102501–102504 (2002).
- Blank, B. *et al.* First observation of ⁵⁴Zn and its decay by two-proton emission. *Phys. Rev. Lett.* **94**, 232501–232504 (2005).
- Schmidt, K. *et al.* Decay properties of the new isotopes ⁹⁴Ag and ⁹⁵Ag. *Z. Phys. A* **350**, 99–103 (1994).
- Mukha, I. *et al.* Observation of direct proton decay of (21⁺) high-spin isomer in ⁹⁴Ag. *Phys. Rev. Lett.* **95**, 022501–022504 (2005).
- Grigorenko, L. V., Johnson, R. C., Mukha, I. G., Thompson, I. J. & Zhukov, M. V. Theory of two-proton radioactivity with application to ¹⁹Mg and ⁴⁸Ni. *Phys. Rev. Lett.* **85**, 22–25 (2000).
- Grigorenko, L. V., Johnson, R. C., Mukha, I. G., Thompson, I. J. & Zhukov, M. V. Two-proton radioactivity and three-body decay. I. General problems and theoretical approach. *Phys. Rev. C* **64**, 054002–054013 (2001).
- Grigorenko, L. V., Mukha, I. G. & Zhukov, M. V. Prospective candidates for the two-proton decay studies. II. Exploratory studies of ³⁰Ar, ³⁴Ca and ⁴⁵Fe. *Nucl. Phys. A* **714**, 425–440 (2003).
- Grigorenko, L. V. & Zhukov, M. V. Two-proton radioactivity and three-body decay. II. Exploratory studies of lifetimes and correlations. *Phys. Rev. C* **68**, 054005–054019 (2003).
- Cerny, J. & Hardy, J. C. Delayed proton radioactivities. *Annu. Rev. Nucl. Part. Sci.* **27**, 333–351 (1977).
- Bochkarev, O. V. *et al.* Democratic decay of ⁶Be states. *Nucl. Phys. A* **505**, 215–240 (1989).
- Kryger, R. A. *et al.* Two-proton emission from the ground state of ¹²O. *Phys. Rev. Lett.* **74**, 860–863 (1995).
- Gomez del Campo, J. *et al.* Decay of a resonance in ¹⁸Ne by the simultaneous emission of two protons. *Phys. Rev. Lett.* **86**, 43–46 (2001).
- Zerguerras, T. *et al.* Study of light proton-rich nuclei by complete kinematics measurements. *Eur. Phys. J. A* **20**, 389–396 (2004).
- Chromik, M. J. *et al.* Two-proton spectroscopy of low-lying states in ¹⁷Ne. *Phys. Rev. C* **66**, 024313–024324 (2002).
- Cable, M. D. J. *et al.* Discovery of beta-delayed two-proton radioactivity: ²²Al. *Phys. Rev. Lett.* **50**, 404–406 (1983).
- La Commara, M. *et al.* Beta decay of medium and high spin isomers in ⁹⁴Ag. *Nucl. Phys. A* **708**, 167–180 (2002).
- Plettner, C. *et al.* On the beta-decaying (21⁺) spin gap isomer in ⁹⁴Ag. *Nucl. Phys. A* **733**, 20–36 (2004).
- Mukha, I. *et al.* Beta-delayed proton decay of a high-spin isomer in ⁹⁴Ag. *Phys. Rev. C* **70**, 044311–044322 (2004).
- Audi, G., Wapstra, A. H. & Thibault, C. The AME2003 atomic mass evaluation (II): Tables, graphs, and references. *Nucl. Phys. A* **729**, 337–676 (2003).
- Kast, D. *et al.* First identification and shell model structure of ⁹²Rh. *Z. Phys.* **356**, 363–365 (1997).
- Ohsen, G. G. Kinematic relations of the form $A + B = C + D + E$. *Nucl. Instrum. Methods* **37**, 240–248 (1965).
- Bochkarev, O. V. *et al.* Three-body decay of 5/2[−] states of ⁹Be and ⁹B isotopic doublet. *Sov. J. Nucl. Phys.* **52**, 964–969 (1990).
- Twin, P. J. *et al.* Observation of discrete-line superdeformed band up to 60(h) in ¹⁵²Dy. *Phys. Rev. Lett.* **57**, 811–814 (1986).
- Phillips, R. J. N. n-p polarization and low energy p-p phase shifts. *Nucl. Phys.* **21**, 686–695 (1960).
- Roedel, E. *et al.* Recent developments at the ISOL facility of GSI Darmstadt. *Nucl. Instrum. Methods Phys. Res. B* **204**, 53–57 (2003).
- Kirchner, R. On the release and ionization efficiency of catcher-ion-source systems in isotope separation on-line. *Nucl. Instrum. Methods Phys. Res. B* **70**, 186–199 (1992).
- Lane, A. M. & Thomas, R. G. R-matrix theory of nuclear reactions. *Rev. Mod. Phys.* **30**, 257–353 (1958).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements Z.J. acknowledges support from a Polish grant. The support of the Université Catholique de Louvain, Belgium, is acknowledged by I.M. L.G. acknowledges support from the Russian RFBR and the Russian Ministry of Industry and Science. The support of the US National Science Foundation is acknowledged by S.L.T.

Author Contributions R.K. was in charge of the development of the ion sources of the on-line mass separator and of the operation of this facility. I.M. suggested and built the charged-particle detector, applied the 'tagging' method in the data analysis and established the reported phenomenon, J.D. set up the γ -ray

detectors and the related electronics and data acquisition. All authors except L.G., M.H. and P.V.D participated in the measurements. L.G. contributed the theoretical interpretation of the data. The manuscript was prepared by I.M. and E.R. with the active participation of all authors, especially J.D., H.G., M.H., S.L.T. and P.V.D.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to I.M. (I.Mukha@gsi.de).

Artificial 'spin ice' in a geometrically frustrated lattice of nanoscale ferromagnetic islands

R. F. Wang¹, C. Nisoli¹, R. S. Freitas¹, J. Li¹, W. McConville¹, B. J. Cooley¹, M. S. Lund², N. Samarth¹, C. Leighton², V. H. Crespi¹ & P. Schiffer¹

Frustration, defined as a competition between interactions such that not all of them can be satisfied, is important in systems ranging from neural networks to structural glasses. Geometrical frustration, which arises from the topology of a well-ordered structure rather than from disorder, has recently become a topic of considerable interest¹. In particular, geometrical frustration among spins in magnetic materials can lead to exotic low-temperature states², including 'spin ice', in which the local moments mimic the frustration of hydrogen ion positions in frozen water^{3–6}. Here we report an artificial geometrically frustrated magnet based on an array of lithographically fabricated single-domain ferromagnetic islands. The islands are arranged such that the dipole interactions create a two-dimensional analogue to spin ice. Images of the magnetic moments of individual elements in this correlated system allow us to study the local accommodation of frustration. We see both ice-like short-range correlations and an absence of long-range correlations, behaviour which is strikingly similar to the low-temperature state of spin ice. These results demonstrate that artificial frustrated magnets can provide an uncharted arena in which the physics of frustration can be directly visualized.

In one of the most common frustrated systems, ordinary water ice, hydrogen ions follow the so-called 'ice rules'. These rules require that the four hydrogens surrounding each oxygen atom be placed in a tetrahedral coordination such that two are close to the central oxygen atom, while the other two are closer to neighbouring oxygen atoms⁷. The spin ice materials have the pyrochlore structure in which magnetic rare-earth ions form a lattice of corner-sharing tetrahedra. In these materials, the ice rules are manifested by a minimization of the spin–spin interaction energy when two spins point inward and two spins point outward on each tetrahedron. At low temperatures, the spins freeze into an exotic disordered state that has many of the hallmarks of glassiness^{3–5}. This ice-like state is different from the disorder-based spin glasses, in that it is associated with a very narrow range of spin relaxation times due to the well-ordered lattice^{8,9}, and it is also quite different from spin liquid states seen in other geometrically frustrated rare-earth magnets^{10,11}, in that the spins do give evidence of freezing into a static configuration at the lowest temperatures.

One of the most fascinating aspects of geometrically frustrated magnets is how the spins locally accommodate the frustration of the spin–spin interactions. As a practical matter, however, individual spins within a material are difficult to probe experimentally without altering the state of the system. One way around this problem is to create a frustrated system in which the individual elements can be directly probed. Toward this end, previous workers have fabricated arrays of superconducting rings or Josephson junctions, in which the interacting moments are trapped flux quanta created by the

application of a magnetic field at low temperatures^{12,13}. A much closer analogy to the frustrated magnetic materials can be provided, however, by arrays of interacting single-domain ferromagnetic islands in which the moments are intrinsic (that is, they do not require the application of an external field). Advances in lithography allow great flexibility in the design of ferromagnetic island arrays, and such arrays have the added advantage of being accessible at room temperature. Studies of lines and rectangular arrays of such ferromagnetic islands show that pairwise dipolar interactions between them can be significant^{14–19}. These results suggest that frustration effects should be important if a lattice of such islands can be fabricated with a frustrated geometry.

We studied frustrated arrays consisting of two-dimensional square lattices of elongated permalloy islands (shown schematically in Fig. 1a) with the long axes of the islands alternating in orientation along the two principal directions of the array lattice (fabrication details are given in the Methods section). We studied arrays with lattice parameters ranging from 320 nm to 880 nm, with a fixed island size (80 nm × 220 nm laterally and 25 nm thick). This size is sufficiently small that the atomic spins were ferromagnetically aligned in a single domain (as demonstrated below), but large enough that the moment configuration was stable at 300 K. The moment of each island was approximately 3×10^7 Bohr magnetons (estimated from the known magnetization of permalloy), and the magnetic field from an island was of the order of 10 Oe at the centre of nearest neighbour islands—leading to an interaction energy of the order of 10^{-19} J, equivalent to 10^4 K, between nearest neighbours (the exact value naturally depends on the lattice spacing). The magnetocrystalline anisotropy of permalloy is effectively zero, so that the shape anisotropy of each island (the self-energy of the island's magnetic moment, which is controlled by its shape) forced its magnetic moment to align along the long axis, thus making the islands effectively Ising-like. Finite element modelling of the islands using the OOMMF code²⁰ indicated that the magnetic fields originating from other islands did not substantially alter this Ising-like behaviour.

The intrinsic frustration on this lattice is similar to that in the 'square ice' model (see ref. 21 and references therein), and can best be seen by considering a vertex at which four islands meet. Nearest-neighbour islands on such a vertex are perpendicular to each other, while the second nearest neighbour islands are collinear and directly across the vertex from each other (see Fig. 1b). A pair of moments on a vertex can be aligned either to maximize or to minimize the dipole interaction energy of the pair. As shown in the figure, it is energetically favourable when the moments of the pair of islands are aligned so that one is pointing into the centre of the vertex and the other is pointing out, while it is energetically unfavourable when both moments are pointing inward or both are pointing outward. For

¹Department of Physics and Materials Research Institute, Pennsylvania State University, University Park, Pennsylvania 16802, USA. ²Department of Chemical Engineering and Materials Science, University of Minnesota, Minneapolis, Minnesota 55455, USA.

the vertex as a whole, there are four distinct topologies for the configurations of the four moments with a total multiplicity of 16, as shown in Fig. 1c. We label the configurations I–IV in the order of increasing magnetostatic energy, but no configuration can minimize all of the dipole–dipole interactions (even type I only minimizes the energy for four of the six pairs in a vertex), and thus the system is frustrated.

The lowest energy vertex configurations (I and II) have two of the moments pointing in toward the centre of the vertex, and two pointing out. Although the interactions between all pairs of spins on the vertex are not equivalent, these energetics are analogous to the two-in/two-out ice rules for the atomic moments on a tetrahedron in spin ice. For arrays with a lattice constant of 320 nm, the energy difference between vertices of types I and III is more than twice as large as the energy difference between vertices of types I and II, and the energy difference between types I and IV is more than six times as large (based on OOMMF calculations of relaxed magnetostatic energies). The two-in/two-out motifs (types I and II) therefore dominate within a large manifold of closely spaced low-energy magnetic states. Topological considerations further favour the creation of magnetic states that are dominated by frustrated mixtures of types I and II. For example, a domain boundary between regions of types I and II is essentially seamless, requiring no vertices of types III or IV. The situation contrasts sharply with that of a traditional Ising ferromagnet or antiferromagnet, wherein magnetic domain walls contain highly unfavourable anti-aligned spin pairs.

Magnetic force microscopy (MFM) allowed us to image the orientations of all of the moments in a large area ($10\ \mu\text{m} \times 10\ \mu\text{m}$), far from the edges of the arrays. To enable the system to settle into a low energy configuration, we followed a protocol developed by previous authors^{16,18} and rotated the samples in a magnetic field which decreased stepwise from above to below the coercive field. MFM images of the system after such field treatment revealed no measurable residual magnetic moment for the array, and a ten-fold variation of the step dwell times did not significantly alter the distribution of vertex types described below.

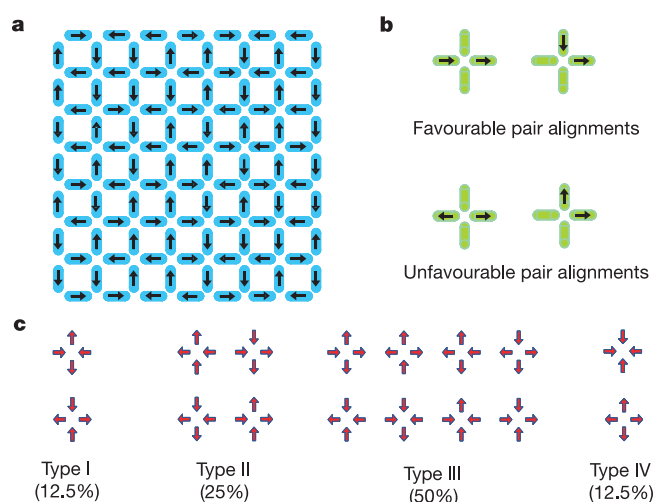


Figure 1 | Illustration of frustration on the square lattice used in these experiments. Each island in the lattice is a single-domain ferromagnet with its moment pointing along the long axis, as indicated by the arrow. **a**, The geometry of the lattice studied. The arrows indicate the directions of moments corresponding to the MFM image of Fig. 2b. **b**, Vertices of the lattice with pairs of moments indicated, illustrating energetically favourable and unfavourable dipole interactions between the pairs. **c**, The 16 possible moment configurations on a vertex of four islands, separated into four topological types. The percentages indicate the expected fraction of each type if the individual moment orientations on an array were completely random.

In Fig. 2 we show an atomic force microscope (AFM) image and an MFM image of a portion of a typical array. The black and white spots in the MFM image, which indicate the north and south poles of the ferromagnetic islands, confirm the single-domain nature of the islands and demonstrate the dominance of the shape anisotropy in aligning the magnetization of each island along its long axis. From the MFM data, we can easily determine the moment configuration of the array (as indicated by the arrows in Fig. 1a). These data demonstrate that the many vertex types anticipated in Fig. 1c can be directly observed in the actual system. In order to probe the nature of frustration in this system, we studied how the properties varied with the spacing between the islands, counting between 1,000 and 3,000 islands in measurements of 2–4 different arrays for each lattice spacing. This allowed us direct control over the frustrated interactions, something which is not easily attainable in geometrically frustrated magnetic materials.

An immediate question is whether our arrays obeyed the ice rules—that is, did a preponderance of the vertices fall into a two-in/two-out configuration (type I or II)? By simple counting arguments (see Fig. 1c) we can predict the expected distribution of different vertex types if the moments were non-interacting and randomly oriented. One would expect only 37.5% of the vertices to have a two-in/two-out configuration if the orientations were random; an excess of such vertices would indicate that interactions are determining the moment configuration. We compute the excess percentage for each type of vertex, defined as the difference between the percentage observed and that expected for a random distribution. We plot this excess versus lattice spacing in Fig. 3a for each of the four vertex types, as well as for types I and II combined. The excess percentage of vertices with a two-in/two-out configuration (types I and II) was well over 30% for the smallest lattice spacing; in other words, over 70% of all vertices had a spin-ice-like configuration. This excess percentage decreased monotonically with increasing lattice spacing (decreasing interactions), approaching zero for our largest lattice spacing, as would be expected for non-interacting (randomly oriented) moments. In fact, the excess for all vertex types approached zero as the lattice spacing increased, lending credence both to our understanding of the system and to the effectiveness of the rotating-field method in enabling facile local re-orientation of the moments.

To further understand the nature of frustration in this system, we also studied the pairwise correlations between the Ising-like moments of the islands. Defining a correlation function is somewhat complicated by the anisotropic nature of our lattice and that of the dipole interaction. We thus define a set of correlation functions between distinct types of neighbouring pairs. The closest pairing is

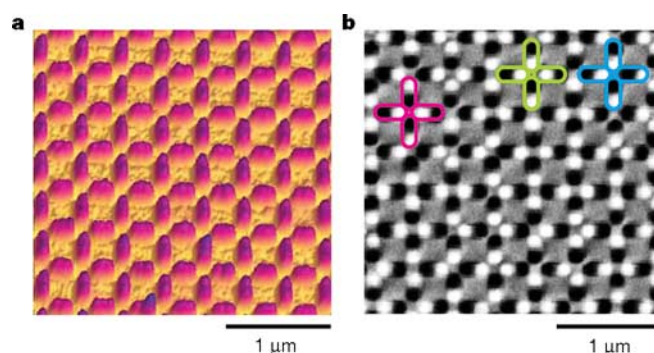


Figure 2 | AFM and MFM images of a frustrated lattice. **a**, An AFM image of a typical permalloy array with lattice spacing of 400 nm. **b**, An MFM image taken from the same array. Note the single-domain character of the islands, as indicated by the division of each island into black and white halves. The moment configuration of the MFM image is illustrated in Fig. 1a. The coloured outlines indicate examples of vertices of types I, II and III in pink, blue and green respectively.

labelled 'NN' for the nearest neighbour; 'L' denotes the next nearest neighbour pairing, which is in the longitudinal direction of the island; and 'T' denotes the next nearest neighbour in the transverse direction from the island (see Fig. 3b inset). We define a correlation, C , such that $C = +1$ if two moments are aligned to minimize the dipole interaction energy, and $C = -1$ if two moments are aligned to maximize the dipole energy. In this way, if the moments for a particular type of neighbouring pair were uncorrelated on the lattice, the average value of C would be zero.

We find that the island pairs which were further separated than the L and T neighbours had weak or zero correlations ($|C| < 0.1$) for all lattices. We do see correlations for the NN and T neighbours as shown in Fig. 3b, but somewhat surprisingly, the correlations for the L neighbours were relatively small. We can understand this as a direct consequence of the frustration in the system. Interaction between the NN neighbours is the strongest, and therefore it is predominant. A pair of islands of type L has a direct interaction (which is somewhat weaker than that of the NN pair), but also an indirect interaction, since the two islands in an L pairing share two NN neighbours. If all of the NN pair energies are minimized, then the L pair energy is maximized, and we believe that this frustration leads to the surprisingly weak correlation between the L neighbours. By the same logic, the relatively strong correlations between the T neighbour pairs also

arise from indirect interactions via NN intermediaries. In the case of the T neighbour pairs, if the NN neighbour pair interaction energy is minimized, the indirect interaction energy will also be minimized, and thus the combined effect is to increase correlations as we observe. For all of the neighbour types, we find that the correlations approached zero for the largest lattice parameters, as expected since the interactions should strongly decrease as the islands are separated.

The existence of only short-range order and ice-like correlations on the lattice is precisely analogous to the behaviour of the spin ice materials, in which there is also no experimental evidence for long-range order, only ice-like short-range correlations. While there are theoretical long-range ordered low-energy states for spins on either our lattice or the pyrochlore spin ice lattice²², the complex energy landscape associated with the frustration leads to a disordered state when thermal or magnetic-field-induced excitations are removed. This is in sharp contrast with unfrustrated lines of ferromagnetic islands, in which longer-range correlations are observed^{16,17}. It is interesting that the relative populations of different types of vertices reaches the randomly oriented limit rapidly as the lattice constant increases, and that even within the regime of closely-spaced and therefore strongly interacting islands, the system can access a very wide range of nearly degenerate states. This wide range of accessible states has the potential for importance to applications, since, if information were encoded within a low energy configuration of the moments, the energetic driving force for local magnetization reversals could be suppressed by this near-degeneracy, even for highly dense arrays.

Our demonstration of an artificial frustrated magnet opens the door to a new mode of research wherein a frustrated system can be designed rather than discovered. Future studies could examine a wide range of accessible lattice geometries, rationally designed defect structures, and the effects of dynamic and static applied magnetic field²³. In addition, the capability to locally probe the magnetic moments, the accessibility at room temperature, and the similarity to patterned magnetic recording media all combine to suggest the potential for novel technological applications that exploit the fundamental nature of frustration.

METHODS

We fabricated the arrays on Si substrates with a native oxide layer, using films of permalloy ($\text{Ni}_{0.81}\text{Fe}_{0.19}$) with grain size of about 5 nm. We employed a lift-off technique using a polymethylglutarimide (PMGI) and polymethyl methacrylate (PMMA) double layer resist^{24,25}. After electron beam exposure, we used methyl isobutyl ketone:isopropyl alcohol (in the ratio of 1:3) to remove exposed PMMA resist, followed by removal of that PMGI resist which is not covered by PMMA. Then we used a molecular beam epitaxy system to grow a 25-nm-thick permalloy film on the pattern at a deposition rate of $0.1 \text{ \AA s}^{-1}$ at ambient temperature. The permalloy was capped with 3 nm of Al to prevent oxidation of the magnetic material. After a lift-off process in acetone, the PMGI and PMMA resists were removed and the nanometre-scale islands stood on the Si substrate. The total array size ranges from $64 \mu\text{m} \times 64 \mu\text{m}$ to $176 \mu\text{m} \times 176 \mu\text{m}$, with the size increasing for less dense arrays (there were 80,000 islands in each array).

We used a Veeco Multimode MFM to detect individual island magnetization under zero magnetic field. Repeated scans demonstrated that the tip did not change the orientation of the island moments. Before measurement, the sample was rotated at 1,000 r.p.m. in an in-plane magnetic field, with the magnetic field starting at 1,300 Oe (well above the coercive field of the islands) and gradually stepping down in magnitude to zero.

Received 16 August; accepted 16 November 2005.

- Ramirez, A. P. in *Handbook of Magnetic Materials* Vol. 13 (ed. Buschow, K. J. H.) 423–520 (Elsevier Science, Amsterdam, 2001).
- Moessner, R. Magnets with strong geometric frustration. *Can. J. Phys.* **79**, 1283–1294 (2001).
- Harris, M. J., Bramwell, S. T., McMorrow, D. F., Zeiske, T. & Godfrey, K. W. Geometrical frustration in the ferromagnetic pyrochlore $\text{Ho}_2\text{Ti}_2\text{O}_7$. *Phys. Rev. Lett.* **79**, 2554–2557 (1997).
- Siddharthan, R. *et al.* Ising pyrochlore magnets: low-temperature properties, "ice rules," and beyond. *Phys. Rev. Lett.* **83**, 1854–1857 (1999).

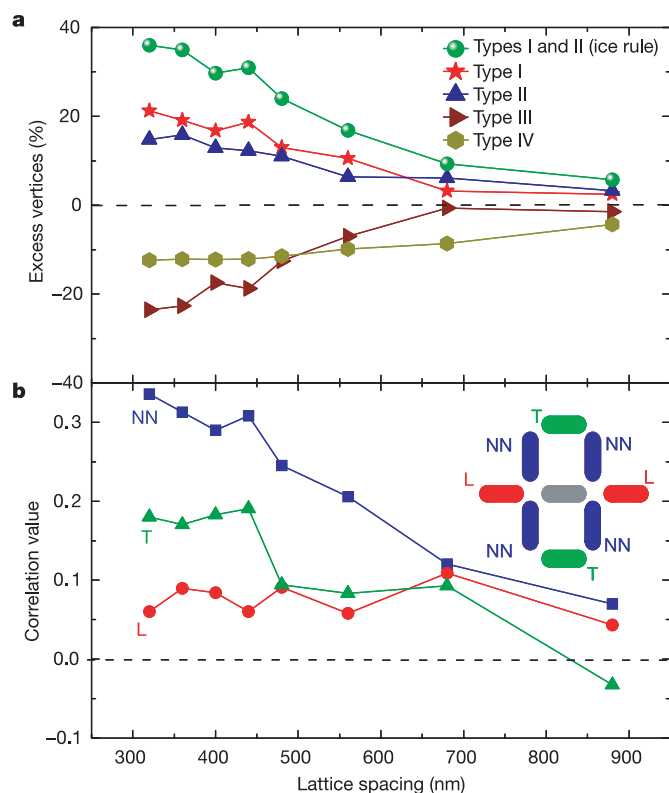


Figure 3 | Statistics of moment configurations. These statistics were obtained from between 1,000 and 3,000 islands in combined measurements of 2–4 different arrays for each lattice spacing. **a**, The excess percentages of different vertex types, plotted as a function of the lattice spacing of the underlying square array lattice. Note that the excess percentages approach zero for the largest lattice spacing. **b**, The correlations between different pairs of the islands as a function of the lattice spacing of the underlying square lattice. The inset shows our definitions of the near neighbour pairs from the grey central island (see text for details). For both the correlations and the vertex statistics, the typical variation between images for nominally identically prepared samples was $<10\%$ for the closely spaced lattices in which we had more than 1,000 islands in a single image, but up to 50% for the more widely spaced lattices in which we had only a few hundred islands per image.

5. Ramirez, A. P., Hayashi, A., Cava, R. J., Siddharthan, R. & Shastry, B. S. Zero-point entropy in 'spin ice'. *Nature* **399**, 333–335 (1999).
6. Bramwell, S. T. & Gingras, M. J. P. Spin ice state in frustrated magnetic pyrochlore materials. *Science* **294**, 1495–1501 (2001).
7. Pauling, L. *The Nature of the Chemical Bond* 301–304 (Cornell Univ. Press, Ithaca, New York, 1945).
8. Snyder, J., Slusky, J. S., Cava, R. J. & Schiffer, P. How 'spin ice' freezes. *Nature* **413**, 48–51 (2001).
9. Snyder, J. *et al.* Low-temperature spin freezing in the $\text{Dy}_2\text{Ti}_2\text{O}_7$ spin ice. *Phys. Rev. B* **69**, 064414 (2004).
10. Tsui, Y. K., Burns, C. A., Snyder, J. & Schiffer, P. Magnetic field induced transitions from spin glass to liquid to long range order in a 3D geometrically frustrated magnet. *Phys. Rev. Lett.* **82**, 3532–3535 (1999).
11. Gardner, J. S. *et al.* Cooperative paramagnetism in the geometrically frustrated pyrochlore antiferromagnet $\text{Tb}_2\text{Ti}_2\text{O}_7$. *Phys. Rev. Lett.* **82**, 1012–1015 (1999).
12. Davidovic, D. *et al.* Correlations and disorder in arrays of magnetically coupled superconducting rings. *Phys. Rev. Lett.* **76**, 815–818 (1996).
13. Hilgenkamp, H. *et al.* Ordering and manipulation of the magnetic moments in large-scale superconducting π -loop arrays. *Nature* **422**, 50–53 (2003).
14. Cowburn, R. P. & Welland, M. E. Room temperature magnetic quantum cellular automata. *Science* **287**, 1466–1468 (2000).
15. Ross, C. A. *et al.* Magnetic behaviour of lithographically patterned particle arrays (invited). *J. Appl. Phys.* **91**, 6848–6853 (2002).
16. Cowburn, R. P. Probing antiferromagnetic coupling between nanomagnets. *Phys. Rev. B* **65**, 092409 (2002).
17. Martin, J. I., Nogues, J., Liu, K., Vicent, J. L. & Schuller, I. K. Ordered magnetic nanostructures: fabrication and properties. *J. Magn. Magn. Mater.* **256**, 449–501 (2003).
18. Imrea, A., Csabaa, G., Bernstein, G. H., Porod, W. & Metlushko, V. Investigation of shape-dependent switching of coupled nanomagnets. *Superlatt. Microstruct.* **34**, 513–518 (2003).
19. Stamps, R. L. & Camley, R. E. Frustration and finite size effects of magnetic dot arrays. *J. Magn. Magn. Mater.* **177**, 813–814 (1998).
20. The Object Oriented MicroMagnetic Framework (OOMMF) project at ITL/NIST. (<http://math.nist.gov/oommf/>) (2005).
21. Lieb, E. H. & Wu, F. Y. in *Phase Transitions and Critical Phenomena* (eds Domb, C. & Green, M. S.) Vol. I (Academic, London, 1972).
22. Melko, R. G., den Hertog, B. C. & Gingras, M. J. P. Long-range order at low temperatures in dipolar spin ice. *Phys. Rev. Lett.* **87**, 067203 (2001).
23. Zhitomirsky, M. E., Honecker, A. & Petrenko, O. A. Field induced ordering in highly frustrated antiferromagnets. *Phys. Rev. Lett.* **85**, 3269–3272 (2001).
24. Yang, X. M. *et al.* Fabrication of sub-50 nm critical feature for magnetic recording device using electron-beam lithography. *J. Vac. Sci. Technol. B* **21**, 3017–3020 (2003).
25. Lin, C. K., Wang, W., Hwu, H. & Chan, Y. Single step electron-beam lithography for asymmetric recess and gamma gate in high electron mobility transistor fabrication. *J. Vac. Sci. Technol. B* **22**, 1723–1726 (2004).

Acknowledgements We acknowledge financial support from the Army Research Office and the National Science Foundation MRSEC programme, and discussions with P. Crowell and P. Lammert. R.S.F. thanks the CNPq-Brazil for sponsorship.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to P.S. (schiffer@phys.psu.edu).

Dynamical fracture instabilities due to local hyperelasticity at crack tips

Markus J. Buehler^{1*} & Huajian Gao^{2*}

As the speed of a crack propagating through a brittle material increases, a dynamical instability leads to an increased roughening of the fracture surface. Cracks moving at low speeds create atomically flat mirror-like surfaces; at higher speeds, rougher, less reflective ('mist') and finally very rough, irregularly faceted ('hackle') surfaces^{1–5} are formed. The behaviour is observed in many different brittle materials, but the underlying physical principles, though extensively debated, remain unresolved^{1–4}. Most existing theories of fracture^{6–12} assume a linear elastic stress–strain law. However, the relation between stress and strain in real solids is strongly nonlinear due to large deformations near a moving crack tip, a phenomenon referred to as hyperelasticity^{13–17}. Here we use massively parallel large-scale atomistic simulations—employing a simple atomistic material model that allows a systematic transition from linear elastic to strongly nonlinear behaviour—to show that hyperelasticity plays a governing role in the onset of the instability. We report a generalized model that describes the onset of instability as a competition between different mechanisms controlled by the local stress field^{6–8} and local energy flow^{13,14} near the crack tip. Our results indicate that such instabilities are intrinsic to dynamical fracture and they help to explain a range of controversial experimental^{1–5,18} and computational^{19–26} results.

Yoffe's model^{6–8} of a moving crack predicts the occurrence of two symmetric peaks of normal stress on inclined cleavage planes at around 73% of the Rayleigh-wave speed. Gao¹² showed that Yoffe's model^{6–8} is consistent with a criterion of crack kinking into the direction of maximum energy release rate. Eshelby¹⁰ and Freund¹¹ argued that the dynamic energy release rate of a rapidly moving crack makes it possible for the crack to split into multiple branches at a critical speed of about 50% of the Rayleigh-wave speed. Marder and Gross presented an analysis including the discreteness of the atomic lattice², and found instability at a speed similar to other models^{6–8,10,11}. Abraham *et al.*^{19,20} suggested that the onset of instability can be understood from the point of view of reduced local lattice vibration frequencies due to softening at the crack tip, and also discussed the onset of the instability in terms of the secant modulus²⁶. Heizler *et al.*²⁵ investigated the crack tip instability with lattice models⁹ using linear stability analysis of the equations of motion, including dissipation. They observed a strong dependence of the instability speed as a function of smoothness of the atomic interaction, and pointed out deficiencies in Yoffe's picture^{6–8}. Gao^{13,14} attempted to explain the reduced instability speed based on the concept of hyperelasticity within the framework of nonlinear continuum mechanics²⁷ (see Supplementary Notes for background information). Supplementary Fig. 1a visualizes the mirror–mist–hackle transition, and Supplementary Fig. 1b depicts a schematic demonstrating the concept of hyperelasticity.

For the atomistic modelling, using large-scale molecular dynamics simulations²⁸ (see Methods and Supplementary Methods), we model a single crystal under mode I loading with an initial crack as depicted in Fig. 1a. Figure 1b and c depicts force versus atomic separation of the interatomic potential (symbols r , Ξ , $d\phi/dr$ and r_{break} are defined in the Methods section 'Interatomic potential'). Figure 1b shows the force versus separation curve with respect to changes of r_{break} , and Fig. 1c shows the variation in shape when Ξ is varied. For small values of $\Xi \approx 50$, the softening effect is quite large. For large values of $\Xi > 1,000$, the amount of softening close to bond breaking becomes

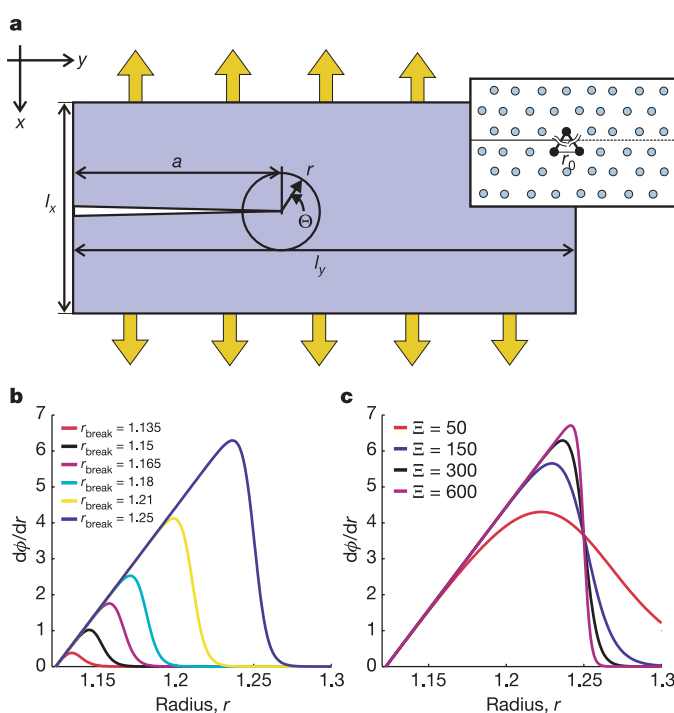


Figure 1 | Simulation geometry with crystal orientation and interatomic force-separation laws. **a**, The simulation geometry with its lattice orientation (l_x and l_y denote the slab sizes in the x and y directions; a is the initial crack length). The interatomic potential showing force versus atomic separation used for the calculations is depicted in **b** and **c** (see also equation (5)). The plots show the change of the shape of the potential depending on its two parameters, r_{break} (in **b**) and Ξ (in **c**). The parameter r_{break} denotes the critical separation for breaking of the atomic bonds and is used to adjust the cohesive stress σ_{coh} . The parameter Ξ corresponds to the amount of smoothing at the breaking. This simplistic model potential leads to generic conclusions about the fracture behaviour of solids.

¹Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Room 1-272, Cambridge, Massachusetts 02139, USA.

²Max Planck Institute for Metals Research, Heisenbergstrasse 3, D-70569 Stuttgart, Germany.

*These authors contributed equally to this work.

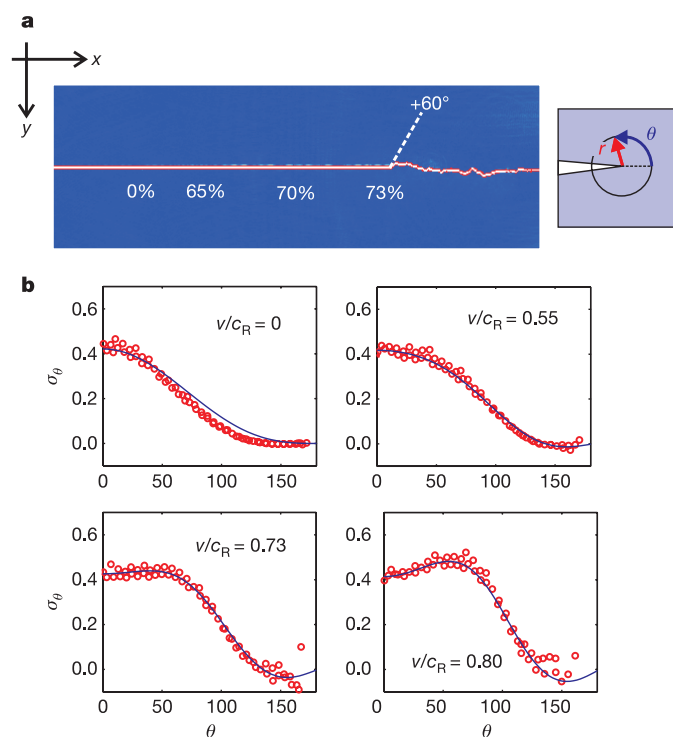


Figure 2 | Crack instability dynamics in the case of harmonic, linear elastic materials behaviour. **a**, The plot shows how the crack starts to branch off at 73% of the Rayleigh-wave speed, in agreement with the classical theories^{6–8}. **b**, Change of circumferential hoop stress σ_θ as a function of crack speed v . The continuous line represents the prediction of continuum theory^{6–8}; the dotted line represents the result of a large-scale molecular dynamics simulation. At 73% of the Rayleigh-wave speed c_R , the hoop stress becomes bimodal, leading into the onset of instability at v_{crit} .

very small, and the solid behaves as though it has snapping bonds^{16,17}. The parameter r_{break} allows the cohesive stress σ_{coh} to be varied independently. This model potential describes the limiting cases of material behaviour corresponding to Yoffe's model^{6–8} (linear elasticity with snapping bonds) and Gao's model^{13,14} (strongly nonlinear behaviour near the crack tip).

Yoffe's model predicts that the instability speed depends only on the small-strain elasticity. Therefore, the instability speed should remain constant at 73% of the Rayleigh-wave speed c_R , regardless of the choices of the parameters r_{break} and Ξ (thus $v_{\text{crit}}^{\text{Yoffe}} \approx 0.73c_R$, where 'crit' stands for 'critical instability'). In contrast, Gao's model predicts that the instability speed $v_{\text{crit}}^{\text{Gao}}$ is dependent on the cohesive stress σ_{coh} and r_{break} (refs 8, 9):

$$v_{\text{crit}}^{\text{Gao}} = \sqrt{\frac{\sigma_{\text{coh}}}{\rho}} \propto \sqrt{\frac{r_{\text{break}}}{\rho}} \quad (1)$$

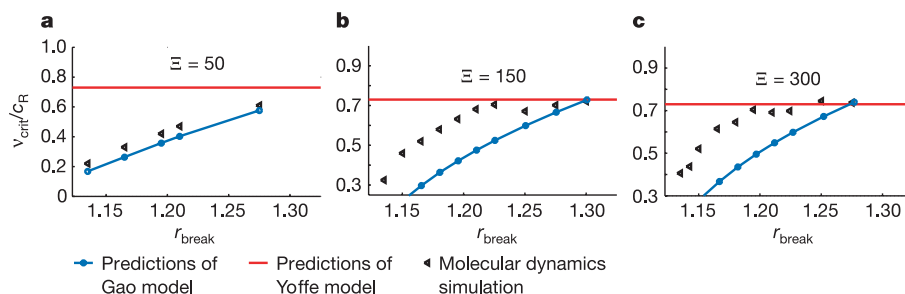


Figure 3 | The critical instability speed as a function of the parameter r_{break} for different choices of the smoothing parameter Ξ . $\Xi = 50$, $\Xi = 150$ and $\Xi = 300$. The results show that the instability crack speed v_{crit} varies with

with material density ρ . Variations in Ξ should not influence the crack instability speed. In contrast to previous work focusing on single potential shapes, here we perform numerical studies based on continuously varying potential parameters r_{break} and Ξ .

To examine whether hyperelasticity governs crack tip instabilities, we carried out a series of numerical experiments by systematically varying the potential parameters r_{break} and Ξ . We start with harmonic systems serving as the reference and then increase the strength of the hyperelastic effect. We find that cracks in homogeneous materials with linear elastic properties (achieved by setting Ξ to infinity) show a critical instability speed of about 73% of the Rayleigh-wave speed, independent of the choice of r_{break} , in quantitative agreement with the prediction by Yoffe's model^{6–8}. The crack surface morphology is shown in Fig. 2a. We find that in harmonic systems, the occurrence of the instability can be correlated with the development of a bimodal hoop stress³⁰ as proposed by Yoffe^{6–8}, as is shown in the sequence of hoop stress as a function of increasing crack speed (Fig. 2b).

Once hyperelastic softening is introduced, we observe that the linear elastic Yoffe model fails to describe the instability dynamics. The predictions of Yoffe's model are included in Fig. 3 as the red line, and the predictions of Gao's model are plotted as the blue points. We observe that for any choice of r_{break} and Ξ , the instability speed lies in between the prediction of Gao's model and that of Yoffe's model. Whether it is closer to Gao's model or to Yoffe's model depends on the choice of r_{break} and Ξ . For small values of Ξ and r_{break} , we find that the instability speed depends on the cohesive stress, which is a feature predicted by Gao's model (equation (1)).

We find that the instability speed seems to be limited by the Yoffe speed (see Fig. 3 for $\Xi = 150$, $\Xi = 300$ and for large values of r_{break}). Whereas the observed limiting speeds increase with r_{break} for $r_{\text{break}} < 1.22$, the results saturate at the Yoffe speed of 73% of Rayleigh-wave speed for $r_{\text{break}} \geq 1.22$ (Fig. 3, bottom curve for $\Xi = 300$). In this case, the instability speed is independent of r_{break} and independent of Ξ . This behaviour suggests a change in mechanism, because the instability speed may be governed by a Yoffe-like deformation field mechanism for $r_{\text{break}} \geq 1.22$, whereas the instability may be influenced by cohesive stress and thus energy flow for $r_{\text{break}} < 1.22$. We observe a similar transition for different choices of Ξ ranging from about 50 to 1,500, with different transition values of r_{break} .

We now present a modified instability model. We observe that the first derivative of the instability speed with respect to the cohesive stress in our molecular dynamics simulations agrees reasonably well with Gao's model. However, the observed instability speed and the prediction differ by a constant value depending on the softening parameter Ξ . We measure the deviation from Gao's model by a shift parameter $v_{\text{shift}} = v_{\text{crit}}^{\text{MD}} - v_{\text{crit}}^{\text{Gao}}$ that is a function of Ξ . The curve is shown in Fig. 4a.

The physical interpretation of v_{shift} is that it accounts for the relative importance of hyperelastic softening close to the crack tip: Gao's model corresponds to the limiting case when the softening region is large and completely dominates the energy flow, and it therefore constitutes the lower limit for the instability speed^{13,14}.

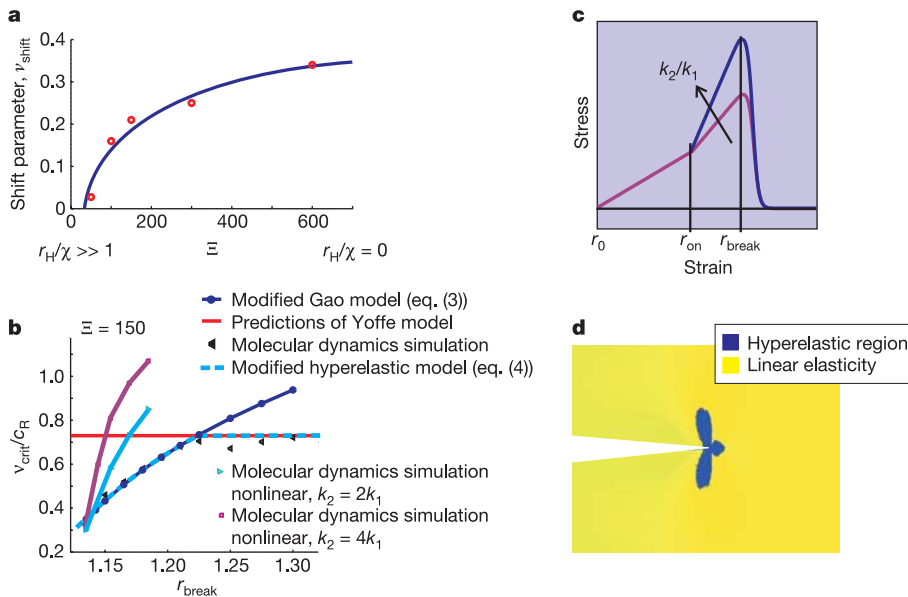


Figure 4 | The modified instability model and stable intersonic crack propagation in stiffening materials. **a**, Change of shift parameter v_{shift} as a function of the smoothing parameter Ξ . **b**, This plot depicts the prediction by the generalized instability model given in equation (4) for $\Xi = 150$ (see upper key in **b**). The plot also contains molecular dynamics simulation results of instability speed for stiffening materials behaviour, clearly showing stable crack motion faster than the Rayleigh-wave speed, as recently observed¹⁹ (see lower key in **b**). **c**, Schematic of stiffening materials behaviour, as assumed in some of the calculations reported in **b**. **d**, Extension of the hyperelastic stiffening region for the case of $r_{\text{break}} = 1.185$, $r_{\text{on}} = 1.1375$ and $k_2/k_1 = 4$. Despite the fact that the stiffening hyperelastic region is highly localized to the crack tip and extends only a few atomic spacings, the crack instability speed is larger than the Rayleigh-wave speed (see **b**). Such stiffening material behaviour is commonly found in rubber-like materials.

Indeed, we find that for very strong softening ($\Xi \rightarrow 0$), v_{shift} vanishes (Fig. 4a). In contrast, v_{shift} assumes larger values in the case of vanishing softening as $\Xi \rightarrow \infty$. The physical significance of this parameter can be understood from the perspective of the characteristic energy length scale of dynamic fracture $\chi \propto \gamma E/\sigma^2$ proposed earlier¹⁶ (see Supplementary Notes for definition of variables). The characteristic energy length scale χ describes the region from which energy flows to the crack tip to drive its motion. The size of the hyperelastic region is defined as r_H . If the size of the softening region is comparable to χ ($r_H/\chi \gg 1$), hyperelasticity dominates energy flow, and thus $v_{\text{shift}} \rightarrow 0$, and the predictions of Gao's model should be valid. In contrast, if the size of the softening region is smaller than χ ($r_H/\chi \approx 0$), hyperelasticity plays a reduced role in the instability dynamics and the purely hyperelastic model becomes increasingly approximate, so that v_{shift} takes larger values and eventually Yoffe's model of deformation-field-induced crack kinking dominates. The shift parameter $v_{\text{shift}}(\Xi)$ should therefore depend on the relative size of the hyperelastic region compared to the characteristic energy flow:

$$v_{\text{shift}} = f\left(\frac{r_H}{\chi}\right) \quad (2)$$

With the new parameter v_{shift} the instability speed is given by:

$$v_{\text{crit}} = v_{\text{shift}}\left(\frac{r_H}{\chi}\right) + \sqrt{\frac{\sigma_{\text{coh}}}{\rho}} \quad (3)$$

Combining these models, we propose a modified instability model in which the critical crack tip instability speed is then given by:

$$v_{\text{crit}}^{\text{mod}} = \min\left(v_{\text{Yoffe}}, v_{\text{shift}}\left(\frac{r_H}{\chi}\right) + \sqrt{\frac{\sigma_{\text{coh}}}{\rho}}\right) \quad (4)$$

where

$$v_{\text{Yoffe}} \approx 0.73c_R \quad (5)$$

is a constant, independent on the hyperelastic properties. Figure 4b compares the predictions from equation (4) (the modified hyperelastic model) with the results of our molecular dynamics experiments (dashed curve).

In the case of stiffening material behaviour, our study reveals stable intersonic mode I crack propagation. This study is motivated by recent computational¹⁶ and experimental¹⁸ observations of intersonic mode I cracks in materials that stiffen with strain. These observations are not explained by existing theories. Our numerical experiments are based on a simple model in which we change the

large-strain and small-strain spring constant¹⁶ (see equation (7) in Methods section 'Interatomic potential'). The model is depicted schematically in Fig. 4c. Upon a critical atomic separation r_{on} (here $r_{\text{on}} = 1.1375$), the spring constant of the harmonic potential is changed and switched to a new 'local' large-strain value (we choose $k_2/k_1 = 2$ and $k_2/k_1 = 4$). Knowledge of r_{on} allows for precise definition of the extension of the hyperelastic zone near the crack tip¹⁶, as can be verified in Fig. 4d.

We observe that if there exists a hyperelastic stiffening zone, the Yoffe speed is no longer a barrier for the instability speed, and stable crack motion beyond the Yoffe speed is possible (Fig. 4b). The stronger the stiffening effect, the more rapid the increase of instability speed with increasing value of r_{break} . All points coincide once r_{break} becomes comparable to r_{on} , corresponding to the case when no hyperelastic stiffening zone is present. The instability speed can even be above the Rayleigh-wave speed and reach intersonic speeds (Fig. 4b for $k_2/k_1 = 4$). This is inconsistent with classical theories^{6–8} and can only be understood from a hyperelasticity point of view^{13,14,16}. The stiffening materials behaviour tends to have a stabilizing effect on straight crack motion owing to enhanced energy flow.

The work reported here together with earlier results¹⁶ strongly suggests that hyperelasticity, including both nonlinear elastic responses and the geometric aspect of large deformation, is crucial for dynamic fracture, both to understand the instability dynamics as well as to comprehend the crack-limiting speed.

The onset of instability can be understood as a competition between energy-flow-governed instability^{13,14} and stress-field-governed instability^{6–8} (equation (4)). We hypothesize that the transition between the two mechanisms depends on the relative importance of hyperelasticity around the crack tip, as described by the ratio of the size of the hyperelastic region and the characteristic energy length scale r_H/χ (ref. 16). In most experiments^{1–5} and computer simulations^{19–26}, materials show a significant softening effect, which explains the reduced instability speed. In the case of a locally stiffening hyperelastic region, the instability speed could exceed the Rayleigh-wave speed. This concept helps to explain recent experimental observations of stable mode I cracking at crack velocities beyond the shear-wave speed¹⁸.

METHODS

Slab geometry, loading condition and typical length scales. The length l_y is four times larger than l_x . The slab width $l_x = 1,150$ in the simulations (corresponding to a few hundred nanometres in physical dimensions). Figure

1a depicts the simulation geometry. All quantities in this paper are given in reduced units. The condition for small-scale yielding is satisfied in all cases (with harmonic, softening and stiffening potentials), because breaking of atomic bonds occurs over a region involving only a few atoms (that is, a very small fracture process zone). There is no dislocation process and the system is perfectly brittle. The crack is oriented such that it initially propagates along the direction of low fracture surface energy (as in several previous studies^{16,17}). In all studies, the slab is under tensile mode I loading (please see Supplementary Methods for further details on the atomistic simulation procedure).

Interatomic potential. Our objective is to develop an interatomic potential that yields elastic behaviour at small and large strains that can be linked to the behaviour of real materials and allows independent variation of parameters governing small-strain and large-strain properties. The model allows us to tune the size of hyperelastic zone and to probe the conditions under which the elasticity of large strains governs the instability dynamics of cracks. Interatomic potentials for a variety of different brittle materials exist, many of which are derived from first principles²⁹. However, it is difficult to identify generic relationships between potential parameters and macroscopic observables (such as the crack instability speed) when using such complicated potentials. We deliberately avoid these complexities by adopting a simple pair potential based on a harmonic interatomic potential with spring constant k_1 in combination with a smooth cut-off of the force based on the Fermi–Dirac distribution function to describe smooth bond breaking. The spring constant k_1 denotes the strength of interatomic bonds. Although simple pair potentials do not allow us to draw conclusions for unusual phenomena pertaining to specific materials, they do enable us to understand universal, generic relationships between potential shape and fracture dynamics in brittle materials; in the present study, we use a simple pair potential that allows the hyperelastic zone size and cohesive stress to be tuned.

A similar approach of using a potential with smoothing parameter has been used in previous studies (albeit slightly different formulation)¹⁶. Unlike in other models²⁵, we do not include any dissipative terms.

The interatomic force $d\phi/dr$ versus atomic separation r is given by:

$$\frac{d\phi}{dr}(r) = k_1(r - r_0) \left[\exp\left(r \frac{\Xi}{r_{\text{break}}} - \Xi\right) + 1 \right]^{-1} \quad (6)$$

where the parameter $r_0 \approx 1.12246$ refers to the nearest-neighbour spacing of atoms. Assuming that the spring constant k_1 is fixed, the potential has two additional parameters, r_{break} and Ξ . The parameter r_{break} (corresponding to the Fermi energy in the Fermi–Dirac function) denotes the critical separation for breaking of the atomic bonds and allows us to tune the breaking strain as well as the cohesive stress σ_{coh} at bond breaking (note that $\sigma_{\text{coh}} \propto d\phi/dr$). The parameter Ξ (corresponding to the temperature in the Fermi–Dirac function) describes the amount of smoothing at the breaking point. We note that similar approaches of developing model potentials for numerical analyses have been done in earlier work¹⁶.

In addition to defining the small-strain elastic properties (by changing the parameter k_1), this model allows us to control the two most critical physical parameters describing hyperelasticity, (1) cohesive stress (by changing the parameter r_{break}), and (2) the strength of softening close to the crack tip (by changing the parameter Ξ).

The bilinear stiffening atomic interaction used above is defined as:

$$\frac{d\phi}{dr}(r) = \left[\exp\left(r \frac{\Xi}{r_{\text{break}}} - \Xi\right) + 1 \right]^{-1} \begin{cases} k_1(r - r_0) & \text{if } r < r_{\text{on}} \\ k_2(r - r_1) & \text{if } r \geq r_{\text{on}} \end{cases} \quad (7)$$

with $r_1 = r_{\text{on}} - \frac{k_1}{k_2}(r_{\text{on}} - r_0)$ from continuation conditions (note that for the special case of $k_1/k_2 = 1/2$ we find $r_1 = \frac{1}{2}(r_{\text{on}} + r_0)$). The parameter k_1 denotes the spring constant associated with small perturbations from the equilibrium distance r_0 , and the second spring constant k_2 is associated with large bond stretching for $r > r_{\text{on}}$, whereas the parameter r_{on} denotes the critical atomic separation at which stiffening sets in. Please see Supplementary Methods for information about calculation of elastic properties.

Crack tip velocity measurements. Accurate determination of crack tip velocity is crucial as we need to be able to measure even the smallest changes in the propagation speed. The crack tip position is determined by finding the surface atom with maximum y position in the interior of a search region inside the slab. This quantity is averaged over a small time interval to eliminate very-high-frequency fluctuations. To obtain the steady-state velocity of the crack, we ensure that the crack dynamics and instability onset is within a region of constant stress intensity factor¹⁰.

Computing techniques and supercomputing centre. A FORTRAN molecular-

dynamics code parallelized with MPI is used for the simulations. The simulations are carried out on IBM Power 4 Regatta nodes at the Max Planck Society Supercomputer Centre in Munich, as well as at the MARS Linux Cluster at the Max Planck Institute for Metals Research in Stuttgart.

Received 23 June; accepted 1 November 2005.

1. Fineberg, J., Gross, S. P., Marder, M. & Swinney, H. L. Instability in dynamic fracture. *Phys. Rev. Lett.* **67**, 141–144 (1991).
2. Marder, M. & Gross, S. Origin of crack-tip instabilities. *J. Mech. Phys. Solids* **43**, 1–48 (1995).
3. Ravi-Chandar, K. Dynamic fracture of nominally brittle materials. *Int. J. Fract.* **90**, 83–102 (1998).
4. Hauch, J. A., Holland, D., Marder, M. P. & Swinney, H. L. Dynamic fracture in single crystal silicon. *Phys. Rev. Lett.* **82**, 3823–3826 (1999).
5. Cramer, T., Wanner, A. & Gumbsch, P. Energy dissipation and path instabilities in dynamic fracture of silicon single crystals. *Phys. Rev. Lett.* **85**, 788–791 (2000).
6. Freund, L. B. *Dynamic Fracture Mechanics* 2nd edn (Cambridge Univ. Press, Cambridge, UK, 1998).
7. Broberg, B. *Cracks and Fracture* (Academic, San Diego, 1999).
8. Yoffe, E. H. The moving Griffith crack. *Phil. Mag.* **42**, 739–750 (1951).
9. Slepian, L. I. *Models and Phenomena in Fracture Mechanics* (Springer, Berlin, 2002).
10. Eshelby, J. D. Fracture mechanics. *Sci. Prog.* **59**, 161–179 (1971).
11. Freund, L. B. Crack propagation in an elastic solid subject to general loading. IV. Obliquely incident stress pulse. *J. Mech. Phys. Solids* **22**, 137–146 (1974).
12. Gao, H. Surface roughening and branching instabilities in dynamic fracture. *J. Mech. Phys. Solids* **41**, 457–486 (1993).
13. Gao, H. A theory of local limiting speed in dynamic fracture. *J. Mech. Phys. Solids* **44**, 1453–1474 (1996).
14. Gao, H. Elastic eaves in a hyperelastic solid near its plane-strain equibiaxial cohesive limit. *Phil. Mag. Lett.* **76**, 307–314 (1997).
15. Guo, G. F., Yang, W. & Huang, Y. Supersonic crack growth in a solid of upturn stress-strain relation under anti-plane shear. *J. Mech. Phys. Solids* **51**, 1971–1985 (2003).
16. Buehler, M. J., Abraham, F. F. & Gao, H. Hyperelasticity governs dynamic fracture at a critical length scale. *Nature* **426**, 141–146 (2003).
17. Buehler, M. J., Abraham, F. F. & Gao, H. in *Multiscale Modeling and Simulation* (eds Attinger, S. & Koumoutsakos, P.) 143–156 (Springer Lecture Notes in Computational Science and Engineering, Springer, Berlin, 2004).
18. Petersen, P. J., Deegan, R. D., Marder, M. & Swinney, H. L. Cracks in rubber under tension exceed the shear wave speed. *Phys. Rev. Lett.* **93**, 015504 (2004).
19. Abraham, F. F., Brodbeck, D., Rudge, W. E. & Xu, X. Instability of fracture—a computer-simulation investigation. *Phys. Rev. Lett.* **73**, 272–275 (1994).
20. Abraham, F. F., Brodbeck, D., Rudge, W. E. & Xu, X. A molecular-dynamics investigation of rapid fracture mechanics. *J. Mech. Phys. Solids* **45**, 1595–1619 (1997).
21. Gumbsch, P., Zhou, S. J. & Holian, B. L. Molecular-dynamics investigation of dynamic crack tip instability. *Phys. Rev. B* **55**, 3445–3455 (1997).
22. Fineberg, J. & Marder, M. Instability in dynamic fracture. *Phys. Rep. Rev. Phys. Lett.* **313**(1–2), 2–108 (1999).
23. Swadener, J. G., Baskes, M. I. & Nastasi, M. Molecular dynamics simulation of brittle fracture in silicon. *Phys. Rev. Lett.* **89**, 085503 (2002).
24. Rountree, C. L. et al. Atomistic aspects of crack propagation in brittle materials: Multimillion atom molecular-dynamics simulations. *Annu. Rev. Mater. Res.* **32**, 377–400 (2002).
25. Heitzler, S. I., Kessler, D. A. & Levine, H. Mode I fracture in a nonlinear lattice with viscoelastic forces. *Phys. Rev. E* **66**, 016126 (2002).
26. Abraham, F. F. Unstable crack motion is predictable. *J. Mech. Phys. Solids* **53**, 1071–1078 (2005).
27. Boresi, A. & Chong, K. P. *Elasticity in Engineering Mechanics* 2nd edn (Wiley-Interscience, New York, 2000).
28. Allen, M. & Tildesley, D. *Computer Simulation of Liquids* (Oxford Univ. Press, New York, 1989).
29. van Duin, A. C. T. et al. ReaxFF_{SiO} reactive force field for silicon and silicon oxide systems. *J. Phys. Chem. A* **107**, 3803–3811 (2003).
30. Zhou, M. A new look at the atomic level virial stress: on continuum-molecular system equivalence. *Proc. R. Soc. A* **459**, 2347–2392 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We acknowledge discussions with F. F. Abraham on atomistic modelling of dynamic fracture.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.J.B. (mbuehler@MIT.EDU).

Low sea level rise projections from mountain glaciers and icecaps under global warming

Sarah C. B. Raper^{1†} & Roger J. Braithwaite²

The mean sea level has been projected to rise in the 21st century as a result of global warming¹. Such projections of sea level change depend on estimated future greenhouse emissions and on differing models, but model-average results from a mid-range scenario (A1B) suggests a 0.387-m rise by 2100 (refs 1, 2). The largest contributions to sea level rise are estimated to come from thermal expansion (0.288 m) and the melting of mountain glaciers and icecaps (0.106 m), with smaller inputs from Greenland (0.024 m) and Antarctica (−0.074 m)¹. Here we apply a melt model³ and a geometric volume model⁴ to our lower estimate of ice volume^{5–7} and assess the contribution of glaciers to sea level rise, excluding those in Greenland and Antarctica. We provide the first separate assessment of melt contributions from mountain glaciers and icecaps, as well as an improved treatment of volume shrinkage. We find that icecaps melt more slowly than mountain glaciers, whose area declines rapidly in the 21st century, making glaciers a limiting source for ice melt. Using two climate models, we project sea level rise due to melting of mountain glaciers and icecaps to be 0.046 and 0.051 m by 2100, about half that of previous projections^{1,8}.

The Intergovernmental Panel on Climate Change (IPCC) estimate^{1,8} takes account of glacier shrinkage under climate warming, but their model uses a time-constant sensitivity for mass balance so that glaciers would melt away completely for any warming rather than approaching a new equilibrium. We apply a glacier mass balance model³ to the total area of glaciers and icecaps⁵, while taking account of changes in glacier area⁴, but our model allows glaciers to approach equilibrium. Our model works on glacier areas within a regular 1° grid instead of irregular regions^{8,9}. The global distribution of mountain glaciers and icecaps^{10,11} that we use is shown in Fig. 1, but we note that the World Glacier Inventory¹² covers only part of the area. The

major challenge for this project is to estimate the parameters that we need in areas not covered by the inventory data. Our glacier areas and volume⁵ do not include mountain glaciers and icecaps around the Greenland and Antarctic ice sheets. We understand that they are nominally included in the 'Greenland' and 'Antarctic' sea level rise (SLR) contributions in ref. 1. Reference 8 uses the same areas as the 100 glacier regions of ref. 9, which clearly exclude Greenland and Antarctica, but assumes a larger ice volume, equivalent to a 0.5-m SLR¹³ that is supposed to include glaciers and icecaps around the ice sheets^{5–7}. The glacier areas and volume in ref. 8 are therefore inconsistent.

We apply our degree-day model³ in regions where we can estimate the average equilibrium line altitude (ELA) from data in the glacier inventory, but we cannot do this for most parts of the world. Our solution is to run the mass balance model for seven regions with good glacier inventory data¹² and then to extrapolate results to the other regions. The seven regions (Axel Heiberg Island, Svalbard, northern Scandinavia, south Norway, the Alps, the Caucasus and New Zealand) cover a wide range of glacial and climatic conditions. We calculate mass balance profiles with the degree-day model for each grid cell within the seven regions and then approximate them with

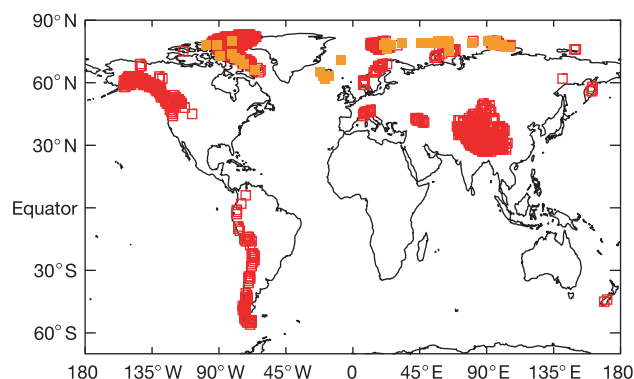


Figure 1 | Worldwide location of grid cells containing glaciers (red)¹⁰ and individual icecaps (orange)¹¹.

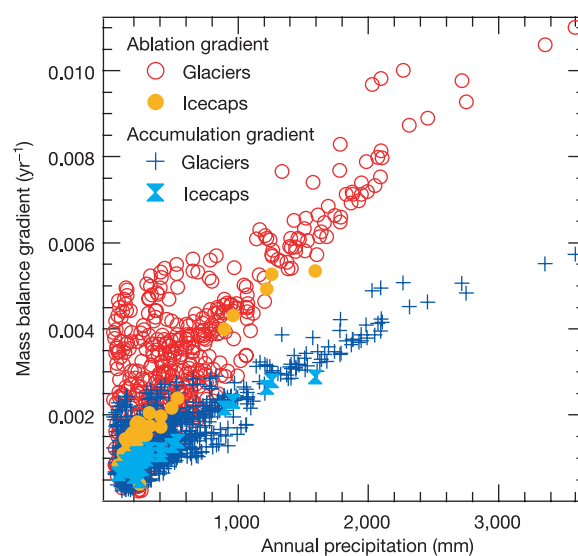


Figure 2 | Altitudinal gradients of mass balance in accumulation and ablation areas plotted against annual precipitation for all grid cells. The values are based on model results for seven regions in which the balance gradient in the accumulation area has correlations of 0.83 and 0.59 respectively with annual precipitation and summer temperature. The corresponding correlations for balance gradient in the ablation area are 0.83 and 0.63.

¹Alfred Wegener Institute for Polar and Marine Research, 27515 Bremerhaven, Germany. ²School of Environment and Development, University of Manchester, Manchester M13 9PL, UK. [†]Present address: CATE, Dalton Research Institute, Manchester Metropolitan University, Manchester M1 5GD, UK.

two segments representing linear gradients of mass balance versus altitude in the accumulation and ablation areas respectively. We regress the resulting balance gradients on annual precipitation and summer temperature from a gridded climatology¹⁴ and we apply the resulting multiple regression equation to all grid cells with glaciers (Fig. 1). There is an obvious association between balance gradients and annual precipitation (Fig. 2): lower balance gradients occur at higher latitudes with cold, dry climate where most icecaps are found. The few larger-gradient values are for the icecaps of Iceland with a relatively warm, wet climate.

We have already estimated^{5–7} size distributions for area and volume of mountain glaciers in 1° latitude/longitude cells as well as area and volume estimates for 116 individual icecaps. Additionally, to run the geometric model⁴, we need the altitude range (minimum to maximum altitude) and area–altitude distribution of the glaciers and icecaps. We estimate the altitudinal range of mountain glaciers from a roughness statistic⁵ derived from high-resolution (30 s) topographic data¹⁵ using a linear regression equation that we calibrate with data from the seven regions where altitude ranges are known. We then approximate the area–altitude distribution of mountain glaciers with a triangle defined by maximum, minimum and mean altitudes. Observed area–altitude distributions for mountain glaciers tend to have a maximum near to the mean altitude where the mass flux of ice is greatest. For an icecap, we assume a parabolic shape with a circular base^{5,16}. This defines both the altitudinal range and the area–altitude distribution, for example, the area within each altitude band increases linearly with altitude up to the top of the icecap.

We calculate mass balance for individual grid squares with mountain glaciers and icecaps and area-weight these to give a global mass balance. We initially assume that the ELA for each mountain glacier or icecap has equal areas above and below, and we then adjust all ELAs to make the model balance fit the estimated global mass balance of $-0.130 \pm 0.033 \text{ m yr}^{-1}$ for the 1961–1990 reference period¹⁷. We note that only a small ELA adjustment ($+18 \pm 17 \text{ m}$) is needed, and this is within the uncertainty limits of observed ELA. This assumed global mass balance value defines the SLR for the reference period as 0.19 mm yr^{-1} .

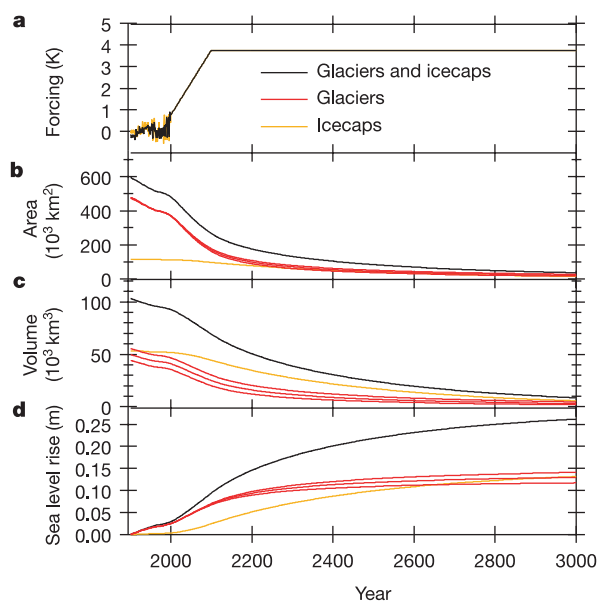


Figure 3 | Time evolution of mountain glacier and icecap metrics. Mountain glacier and icecap area (b), volume (c) and SLR contribution (d). The temperature forcing (a) for the 20th century is from climate data¹⁴ and for the future is an idealized scenario. Results are shown for three reference period mountain glacier volumes (37×10^3 , 43×10^3 and $49 \times 10^3 \text{ km}^3$), but the corresponding total volumes and total SLR contributions are for the central value only. The total potential SLR is 0.26 m.

For comparison with earlier studies, we increase the temperature in the model by 1 K for all months to get a globally averaged mass balance sensitivity of $-0.35 \text{ m yr}^{-1} \text{ K}^{-1}$, which is slightly less than previous estimates of -0.39 , 0.41 and $-0.37 \text{ m yr}^{-1} \text{ K}^{-1}$ (refs 9, 18 and 19, respectively). However, an estimated $\pm 15\%$ uncertainty in our mass balance gradients gives an uncertainty of $\pm 0.050 \text{ m yr}^{-1} \text{ K}^{-1}$ in our mass balance sensitivity, so our new result is not significantly lower than previous estimates. In this experiment, the highest SLR contributions come from individual grid cells in the Gulf of Alaska, Patagonia and Iceland, where we find large mass balance sensitivity to coincide with large ice areas. The large mass balance sensitivity is consistent with the recent high rates of ice loss in Alaska and Patagonia^{20,21} if these regions recently experienced higher temperatures. To reproduce exactly the known glacier wastage region by region would require us to adjust the ELA in the reference period grid point by grid point. This could be the subject of further research. Here we adjust all ELAs uniformly to match the assumed global mean mass balance in the reference period as described above.

Several workers^{1,3,8,9} conclude that the main climate variable controlling past and future global changes in mountain glaciers and icecaps is temperature change, with precipitation being of secondary importance. Reference 1 uses only temperature forcing in the form of anomalies and we do likewise, on the basis of the 1961–1990 reference period. We apply forcing annually by perturbing the ELA from its reference state for each grid cell by the temperature anomaly for each year divided by the lapse rate. (The temperature anomaly is the average over the four summer months in each hemisphere and the assumed lapse rate is 0.006 K m^{-1} .) For each year, the procedure is to perturb the ELA from its reference state, calculate the change in volume from changes in mass balance, and then the resulting changes in total area and area–altitude distribution are calculated with the glacier geometric model⁴.

In the next experiment, we use gridded temperature data (1° latitude by 1° longitude resolution) derived from observations¹⁴ to assess mountain glacier and icecap changes over the 20th century, and we assume uniform warming for 1998–2100, followed by constant temperature to the end of the millennium. The results (Fig. 3) reflect the greater area but smaller volume of the mountain glaciers compared with the icecaps in the reference period⁵. During the 20th century, the areas and volumes for mountain glaciers decline much more than for the icecaps and contribute nearly all the SLR, while icecaps begin to make a significant SLR contribution in the 21st century. The decline in mountain glacier area and volume becomes a limiting factor in the glacier melt contribution to SLR during the

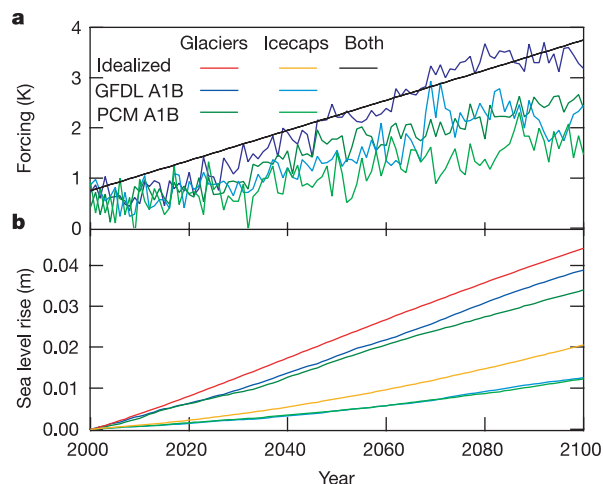


Figure 4 | Temperature forcing and SLR contribution for mountain glaciers and icecaps over the 21st century. The temperature forcing (a) is from two climate models using scenario A1B (ref. 2). The idealized scenario results are shown for comparison. The key in a also refers to the SLR contribution in b.

21st century. Even at the end of the millennium some ice survives, despite prolonged exposure to higher temperature.

We made further experiments to illustrate the glacier and icecap response to spatially differentiated forcing, using results from two climate models, the GFDL and the PCM^{22,23}. The model simulations for GFDL_CM2_0_run1 and for NCAR_PCM1_run2 run through the 20th century, based on historical radiative forcing, followed by the 21st-century mid-range A1B scenario². Figure 4a shows the 21st-century area-averaged mountain glacier and icecap temperature forcing from these two models; the idealized forcing is shown for comparison. The temperature response to radiative forcing of the GFDL model is greater than that of the PCM model, but both models show markedly greater warming over the mountain glacier regions compared with the icecap regions; there is a warming difference of about 1 K by 2100. These temperature differences are reflected in the SLR contributions in Fig. 4b. The main difference between the two projections is the greater mountain glacier melt using the GFDL model compared to PCM. By contrast, icecap melt over the 21st century is very similar for the two simulations.

We summarize here the SLR results for the 20th and 21st centuries (see also the Supplementary Information). On the basis of the observed climate, we estimate the mountain glacier and icecap contribution to SLR to be 0.028 m for the 20th century, which is similar to previous values¹. The GFDL and PCM models give broadly similar SLR values (0.030 and 0.021 m) for the 20th century; for the A1B scenario the models give a much lower SLR for the 21st century (0.046 and 0.051 m) than previous estimates¹. The dominant uncertainty for the 20th century is the uncertainty in area-weighted glacier mass balance for 1961–1990, but its relative importance drops very much in the 21st century compared to high negative mass balances under global warming. For the 21st century, the major uncertainty is the uncertainty in balance gradients, leading to differences in mass balance sensitivity.

Received 1 June; accepted 14 November 2005.

- Church, J. A. *et al.* in *Climate Change 2001: The Scientific Basis* (eds Houghton, J. T. *et al.*) 639–693 (Cambridge Univ. Press, Cambridge, UK, 2001).
- Nakićenović, N. & Swart, R. (eds) *Special Report on Emissions Scenarios* 570 (Cambridge Univ. Press, Cambridge, UK, 2000).
- Braithwaite, R. J., Zhang, Y. & Raper, S. C. B. Temperature sensitivity of the mass balance of mountain glaciers and icecaps as a climatological characteristic. *Z. Gletscherkunde Glazialgeol.* **38**, 35–61 (2002).
- Raper, S. C. B., Brown, O. & Braithwaite, R. J. A geometric glacier model for sea level change calculations. *J. Glaciol.* **46**, 357–368 (2000).
- Raper, S. C. B. & Braithwaite, R. J. The potential for sea level rise: New estimates from glacier and ice cap area and volume distributions. *Geophys. Res. Lett.* **32**, L05502, doi:10.1029/2004GL021981 (2005).
- Meier, F., Bahr, D. B., Dyurgerov, M. B. & Pfeffer, W. T. Comment on 'The potential for sea level rise: New estimates from glacier and ice cap area and volume distributions'. *Geophys. Res. Lett.* **32**, L17501, doi:10.1029/2005GL023319 (2005).
- Raper, S. C. B. & Braithwaite, R. J. Reply to comment by M. F. Meier *et al.* on 'The potential for sea level rise: New estimates from glacier and ice cap area and volume distributions'. *Geophys. Res. Lett.* **32**, L17502, doi:10.1029/2005GL023460 (2005).
- Van de Wal, R. S. W. & Wild, M. Modelling the response of glaciers to climate change by applying the volume-area scaling in combination with a high resolution GCM. *Clim. Dyn.* **18**, 359–366 (2001).
- Oerlemans, J. in *Ice in the Climate System* (ed. Peltier, W. R.) 101–116 (Springer, Berlin/Heidelberg, 1993).
- Cogley, J. G. *GGHYDRO—Global Hydrographic Data Release 2.3*, Trent Technical Note 2003–1 (Department of Geography, Trent Univ., Peterborough, Ontario, 2003); available at (<http://www.trentu.ca/geography/glaciology/glaciology.htm>).
- Kotlyakov, V. M. *et al.* *World Atlas of Snow and Ice Resources* 1–392 (Russian Academy of Sciences, Institute of Geography, Moscow, 1997).
- National Snow and Ice Data Center (NSIDC). *World Glacier Inventory* (http://nsidc.org/data/glacier_inventory/) (NSIDC, Boulder, Colorado, 1999).
- Meier, M. F. & Bahr, D. B. in *Glaciers, Ice Sheets and Volcanoes: A Tribute to Mark F. Meier* (ed. Colbeck, S. C.) 89–94 (Cold Regions Research Engineering Laboratory Special Report 96–27, Hanover, New Hampshire, 1996); available at (<http://www.stormingmedia.us/24/2431/A243123.html>).
- New, M., Hulme, M. & Jones, P. J. Representing twentieth century space-time climate variability. 1. Development of a 1961–1990 mean monthly terrestrial climatology. *J. Clim.* **12**, 829–856 (1999).
- National Geodetic Survey data set. *GLOBE*. Available at the National Geophysical Data Center (<http://www.ngdc.noaa.gov/mgg/topo/globe.html>) (NGS, 1997).
- Paterson, W. S. B. *The Physics of Glaciers* 480 (Pergamon, Oxford, 1994).
- Dyurgerov, M. B. & Meier, M. F. Mass balance of mountain and subpolar glaciers: a new assessment for 1961–1990. *Arct. Alp. Res.* **29**, 379–391 (1997).
- Braithwaite, R. J. & Raper, S. C. B. Glaciers and their contribution to sea level change. *Phys. Chem. Earth* **27**, 1445–1454 (2002).
- Dyurgerov, M. B. & Meier, M. F. Twentieth century climate change: evidence from small glaciers. *Proc. Natl Acad. Sci. USA* **97**, 1406–1411 (2000).
- Arendt, A. A., Echelmeyer, K. A., Harrison, W. D., Lingle, C. S. & Valentine, V. B. Rapid wastage of Alaska glaciers and their contribution to rising sea level. *Science* **297**, 382–386 (2002).
- Rignot, E., Rivera, A. & Casassa, G. Contribution of the Patagonia Icefields of South America to sea level rise. *Science* **302**, 434–437 (2003).
- Meehl, G. A. *et al.* How much more global warming and sea level rise? *Science* **307**, 1769–1772 (2005).
- Geophysical Fluid Dynamics Laboratory. *GFDL CM2 Model* (<http://data1.gfdl.noaa.gov/nomads/forms/deccen/>) (GFDL, 2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements Much of this work was supported by the HGF Strategiefonds Projekt (S.C.B.R.). We acknowledge the international modelling groups for providing their data for analysis, the Program for Climate Model Diagnosis and Intercomparison (PCMDI) for collecting and archiving the model data, the JSC/CLIVAR Working Group on Coupled Modelling (WGCM) and their Coupled Model Intercomparison Project (CMIP) and Climate Simulation Panel for organizing the model data analysis activity, and the IPCC WG1 TSU for technical support. The IPCC Data Archive at Lawrence Livermore National Laboratory is supported by the Office of Science, US Department of Energy.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to S.C.B.R. (s.raper@mmu.ac.uk).

LETTERS

Thallium isotopic evidence for ferromanganese sediments in the mantle source of Hawaiian basalts

Sune G. Nielsen^{1,2}, Mark Rehkämper^{1,3}, Marc D. Norman⁴, Alex N. Halliday^{1,5} & Darrell Harrison¹

Ocean island basalts are generally thought to be the surface expression of mantle plumes, but the nature of the components in the source regions of such mantle plumes is a subject of long-standing debate. The lavas erupted at Hawaii have attracted particular attention, as it has been proposed that coupled ^{186}Os and ^{187}Os anomalies reflect interaction with the Earth's metallic core^{1,2}. It has recently been suggested, however, that such variations could also result from addition of oceanic ferromanganese sediments to the mantle source of these lavas^{3–5}. Here we show that Hawaiian picrites with osmium isotope anomalies also exhibit pronounced thallium isotope variations, which are coupled with caesium/thallium ratios that extend to values much lower than commonly observed for mantle-derived rocks. This correlation cannot be created by admixing of core material, and is best explained by the addition of ferromanganese sediments into the Hawaii mantle source region. However, the lack of correlation between thallium and osmium isotopes and the high thallium/osmium ratios of ferromanganese sediments preclude a sedimentary origin for the osmium isotope anomalies, and leaves core–mantle interaction as a viable explanation for the osmium isotope variations of the Hawaiian picrites.

The isotopic compositions of ocean island basalts (OIB) are easily resolvable from those of the mantle-derived basalts erupted at mid-ocean ridges (mid-ocean-ridge basalts, MORB), and therefore it is clear that the sources of these types of magmatism must also be different. One model infers that OIB originate from a deep mantle source, which was previously contaminated with ocean crust that entered the mantle via subduction zones⁶. On the basis of $^{187}\text{Os}/^{188}\text{Os}$ and $^{186}\text{Os}/^{188}\text{Os}$ isotopic evidence, it was proposed that the Earth's core may also contribute up to 0.5–1% by weight to some mantle plumes^{1,7–9}. These arguments have been supported by the elevated Fe/Mn ratios of Hawaiian picrites that also exhibit Os isotope anomalies¹⁰. However, the lack of tungsten isotope anomalies in some of the same samples⁴ may exclude a core source, though it has been suggested that W, Os and Fe/Mn ratios could be decoupled during core–mantle reaction processes¹⁰. As an alternative to core–mantle interaction, it was noted by several authors that oceanic ferromanganese (Fe–Mn) sediments display very high Pt/Os, such that they would develop high $^{186}\text{Os}/^{188}\text{Os}$ over time^{3–5}. Addition of ancient Fe–Mn sediments into a mantle source, therefore, could also explain the elevated $^{186}\text{Os}/^{188}\text{Os}$ ratios observed for some mantle plumes.

As pointed out in ref. 3, additions of Fe–Mn sediments to a mantle source should generate ^{186}Os isotope anomalies that are accompanied by significant Tl isotope variations. Such a co-variation is expected because Fe–Mn sediments have high Tl concentrations and fractionated Tl isotope compositions with $\epsilon^{205}\text{Tl}$ values as high as +15 (refs 11, 12). Moreover, recent studies have shown that the

continental crust and upper mantle are characterized by relatively constant Tl isotope ratios of $\epsilon^{205}\text{Tl} = -2.0 \pm 0.5$ (refs 13, 14). This indicates that the Tl isotope system may be an excellent monitor of Fe–Mn sediment additions to OIB source regions. We have therefore measured the Tl abundances and isotope compositions of nine Hawaiian picrites that have previously been investigated for ^{187}Os – ^{186}Os isotope systematics¹ and two additional picrites from Mauna Kea and Kilauea¹⁵.

The Tl isotope compositions of the Hawaiian picrites vary in $\epsilon^{205}\text{Tl}$ from -3.1 to $+3.8$ (Table 1). The most positive values are about 6 ϵ -units 'heavier' than two Icelandic basalts (Table 1) and the depleted mantle as represented by MORB (Table 1), high-temperature hydrothermal fluids and sheeted dykes from Ocean Drilling Program Hole 504B¹³. Additionally, the Tl isotope data display a negative correlation with Cs/Tl ratios (Fig. 1) and Rb/Tl (not plotted). Thallium and Cs (and Rb) exhibit very similar incompatibilities in igneous processes¹⁶ and are therefore not expected to fractionate significantly during partial melting or magmatic differentiation. As a consequence, it appears likely that a component with positive $\epsilon^{205}\text{Tl}$ and low Cs/Tl contributed to the Hawaiian lavas. In addition to high Tl contents and positive $\epsilon^{205}\text{Tl}$ -values, Fe–Mn sediments are known to display low Cs/Tl ratios¹⁷. The correlation of Fig. 1 is therefore most readily explained by the addition of Fe–Mn sediments into the mantle source region of Hawaiian magmatism.

There are other processes that could generate lavas with positive $\epsilon^{205}\text{Tl}$ -values. Thallium is a highly volatile element, and magma degassing, which is probably a significant factor controlling the abundances of the elements Cd, Bi and Re in Hawaiian basalts¹⁸, could conceivably cause kinetic Tl isotope fractionation. This process would be expected to produce residual degassed lavas with low Tl contents and positive $\epsilon^{205}\text{Tl}$ -values. However, as volcanic exhalations display low Cs/Tl ratios of less than 0.5 (ref. 19), increases in $\epsilon^{205}\text{Tl}$ would be associated with increasing Cs/Tl ratios, which is the opposite of the observed trend (Fig. 1). Degassing can therefore be excluded as the cause of the observed Tl isotope variations.

A recent study has shown that the short-lived radionuclide ^{205}Pb , which decays to ^{205}Tl , was present in the early Solar System²⁰ and therefore probably also at the time of terrestrial core formation. A planetary core with a high Pb/Tl ratio could thus develop a significant Tl isotope anomaly. Moreover, Cs is a highly lithophile element whereas Tl is mildly chalcophile, such that the core is expected to have a low Cs/Tl ratio. If the correlation of Fig. 1 is to be explained by binary core–mantle mixing, this necessitates a core with Cs/Tl ≈ 0 to 2 coupled with $\epsilon^{205}\text{Tl} \approx +4$ to $+6$. The Tl content of the core for a given Tl isotope ratio can be estimated, by calculating the Tl and Pb isotope evolution for terrestrial accretion models very similar to those used of ref. 21. Owing to the low initial Solar System abundance

¹Department of Earth Sciences, ETH Zurich, Sonneggstrasse 5, 8092 Zurich, Switzerland. ²GEMOC, Department of Earth and Planetary Sciences, Macquarie University, 2109 New South Wales, Australia. ³Department of Earth Science and Engineering, Imperial College, London SW7 2AZ, UK. ⁴Research School of Earth Sciences, Australian National University, Canberra, Australian Capital Territory 0200, Australia. ⁵Department of Earth Sciences, University of Oxford, Parks Road, Oxford OX1 3PR, UK.

Table 1 | Tl isotope compositions and concentrations for Hawaii picrites

Sample	Location	$\epsilon^{205}\text{Tl}$	Tl (ng g^{-1})	Cs (ng g^{-1})†	Rb ($\mu\text{g g}^{-1}$)†	Cs/Tl
Kil 1-18	Kilauea	-3.1	16.9	78	6.3	4.6
LO 02-02	Loihi	-2.0	12.2	77	6.5	6.3
ML 1868-9*	Mauna Loa	-1.4	8.5	43	4.3	5.0
Kil 93-1489*	Kilauea	-1.2	14.7	74	7.2	5.0
Kil 1840*	Kilauea	-1.2	10.9	65	7.0	5.9
MK 1-6	Mauna Kea	-0.6	19.8	75	6.9	3.8
Kil 1-7	Kilauea	-0.6	18‡	71	5.8	3.9
KOO-CF*	Ko'olau	-0.5	22.9	88	7.6	3.8
H-11	Hualalai	0.9	13.5	44	4.5	3.3
KOO-17a*	Ko'olau	1.3	1.7	5	0.3	2.9
ML-2-50	Mauna Loa	3.8	15.1	35	3.4	2.3
RSG 19*	Iceland	-2.0	ND	ND	ND	ND
SNS 14*	Iceland	-1.9	ND	ND	ND	ND
SO157 54DS1	Pacific-Antarctic rise	-1.6	7.4	ND	ND	ND
POS 221 626 DS	Kolbeinsey ridge	-2.5	7.1	ND	ND	ND
TT 152-21	Juan de Fuca ridge	-2.5	9.7	ND	ND	ND
ALV 731-4	Galapagos ridge	-1.8	5.5	ND	ND	ND
CH 98 DR11	Mid-Atlantic Ridge	-0.9	4.7	ND	ND	ND

$\epsilon^{205}\text{Tl} = \{[(^{205}\text{Tl}/^{203}\text{Tl})_{\text{meas}} - (^{205}\text{Tl}/^{203}\text{Tl})_{\text{std}}]/(^{205}\text{Tl}/^{203}\text{Tl})_{\text{std}}\} \times 10^4$, where $(^{205}\text{Tl}/^{203}\text{Tl})_{\text{std}}$ is the Tl isotope composition of the NIST 997 Tl isotope standard. ND, not determined.

*Subaerial samples.

†Cs and Rb concentrations from ref. 15, except Kil 93-1489 from ref. 30.

‡For this sample, the Tl concentration has an uncertainty of $\pm 50\%$, owing to loss of sample during dissolution.

of ^{205}Pb ($^{205}\text{Pb}/^{204}\text{Pb} \approx 1.5 \times 10^{-4}$, ref. 20) and the prolonged accretion of the Earth (99% accreted at ~ 45 Myr after Solar System formation²¹), Tl isotope anomalies in the core only reach sufficient levels (that is, $\epsilon^{205}\text{Tl} > +4$) at very high Pb/Tl ratios of > 240 . The following calculation illustrates the maximum attainable Tl concentration for a core with $\epsilon^{205}\text{Tl} \approx +4$, which is the minimum value required to explain the most anomalous Hawaiian sample. The Pb content of the bulk Earth is unknown, but a reasonable upper limit is set by the solar Pb abundance (as defined by CI chondrites), which is about $2,400 \text{ ng g}^{-1}$ (ref. 22). As the silicate portion of the Earth contains about 150 ng g^{-1} Pb (ref. 23), mass balance dictates an upper limit to the Pb concentration in the core of $\sim 7,300 \text{ ng g}^{-1}$. Combining this Pb abundance with a Pb/Tl ratio of 240 yields a Tl content of approximately 30 ng g^{-1} . Assuming that the lower mantle contains 0.5 ng g^{-1} Tl (a low estimate compared to 3 ng g^{-1} estimated in ref. 23), entrainment of more than 10% core material is required to obtain a mantle source with $\epsilon^{205}\text{Tl} > +3.2$, comparable to the largest isotope anomaly found in Hawaii (Table 1). We note that faster accretion rates would lower the needed Pb/Tl (for example, 99% accretion at 30 Myr necessitates $\text{Pb/Tl} > 200$, resulting in a core Tl concentration of $\sim 40 \text{ ng g}^{-1}$). However, this will only reduce the amount of entrained core material to $\sim 8\%$. The quantities of core material required are more than an order of magnitude larger than the most recent estimates based on ^{186}Os anomalies⁷, and inconsistent with the highly siderophile element abundances and Pb isotope compositions of these lavas^{15,24}. Thus, we conclude that the core is not a viable source for the Tl isotope anomalies of mantle plumes.

The above arguments demonstrate that the observed correlation between $\epsilon^{205}\text{Tl}$ and Cs/Tl (Fig. 1) is most readily explained by additions of marine Fe-Mn precipitates. We note that the precipitation of Fe-Mn oxyhydroxides onto submarine basalts does not provide a feasible explanation for the Hawaiian Tl data, as subaerial samples also display Tl isotope effects (Table 1). Moreover, the negative correlation of Tl and Pb isotopes towards unradiogenic values (excluding samples from Ko'olau) (Fig. 2) attests to the ancient nature of the Fe-Mn component and renders assimilation of recent ferromanganese sediments during magma ascent as a cause for the Tl isotope anomalies unlikely. We conclude that the most reasonable process for generating the observed Tl isotopic compositions of Hawaiian basalts is contamination of their mantle source by Fe-Mn sediments. Quantitative modelling of the mixing process is made difficult, however, by possible temporal variations in the Tl isotope composition of sea water and hence of Fe-Mn precipitates¹².

If it is assumed that the Fe-Mn sediments are characterized by $\epsilon^{205}\text{Tl} = +10$ (ref. 11), akin to the average value of modern Fe-Mn crusts and nodules, then less than $20 \mu\text{g g}^{-1}$ (of Fe-Mn sediments by weight) need to be admixed to a pristine mantle source to account for the Tl isotope variations of the Hawaiian picrites. The slope of the mixing line that is obtained for this scenario is slightly shallower than the actual data array, however (Fig. 1). A best-fit line through the

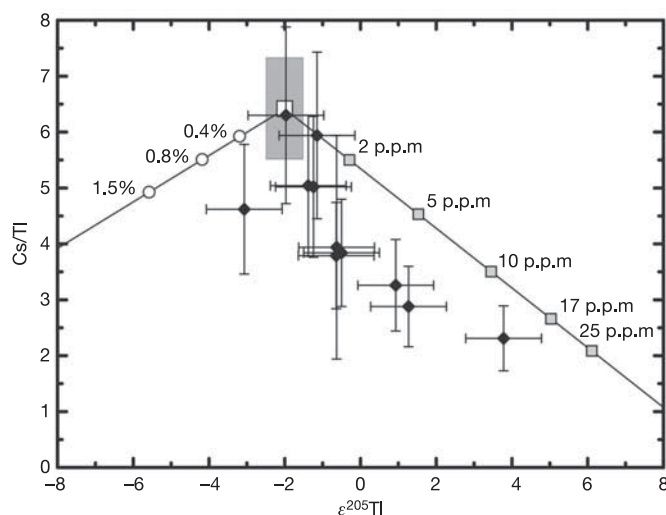


Figure 1 | Thallium isotope compositions of the Hawaiian picrites plotted versus Cs/Tl ratios. Mixing lines between pristine mantle (large open square), Fe-Mn oxyhydroxides (small grey squares) and altered MORB (open circles) are also plotted. The amounts (by weight) mixed into the mantle source are labelled on the mixing lines. For the mantle, the Tl isotope composition is $\epsilon^{205}\text{Tl} = -2.0$ (ref. 13), and a Cs concentration of 7.7 ng g^{-1} , is used³¹. The mantle Tl concentration is estimated to be about 1.2 ng g^{-1} , from $\text{Cs/Tl} = 6.5$ for the Hawaiian sample with the most mantle-like Tl isotope composition (LO 02-02). This Cs/Tl ratio is also consistent with a previous estimate of the mantle's average Cs/Tl ratio³². For the Fe-Mn oxyhydroxides, the Tl concentration and isotope composition are assumed to be $100 \mu\text{g g}^{-1}$ and $\epsilon^{205}\text{Tl} = +10$, akin to values of modern Fe-Mn crusts and nodules^{11,17}. The Cs content of Fe-Mn oxyhydroxides is about 500 ng g^{-1} (ref. 33). Altered MORB is assumed to be characterized by $\epsilon^{205}\text{Tl} = -15$ (ref. 13) and Tl and Cs concentrations of 30 ng g^{-1} (ref. 13). Error bars denote 2 s.d. uncertainties. The large grey shaded rectangle denotes the range of the mantle Cs/Tl and $\epsilon^{205}\text{Tl}$ values.

data would imply that the Fe-Mn component is characterized by $\varepsilon^{205}\text{Tl} \approx +5$, which is not unrealistic¹². Alternatively, the deviations of the picrite data from the calculated binary mixing line of Fig. 1 may reflect the additional presence of small amounts (less than about 1%) of recycled low-temperature altered MORB, as was previously suggested on the basis of $^{187}\text{Os}/^{188}\text{Os}$ and oxygen isotopic evidence²⁵. The scatter of the picrite data in a plot of $^{206}\text{Pb}/^{204}\text{Pb}$ versus $\varepsilon^{205}\text{Tl}$ can also be readily accommodated, by assuming that the Pb isotope compositions of the samples are dominated by variable additions of Fe-Mn sediments and an additional lithogenic marine sedimentary component, which has a composition that is similar to the upper continental crust (Fig. 2a).

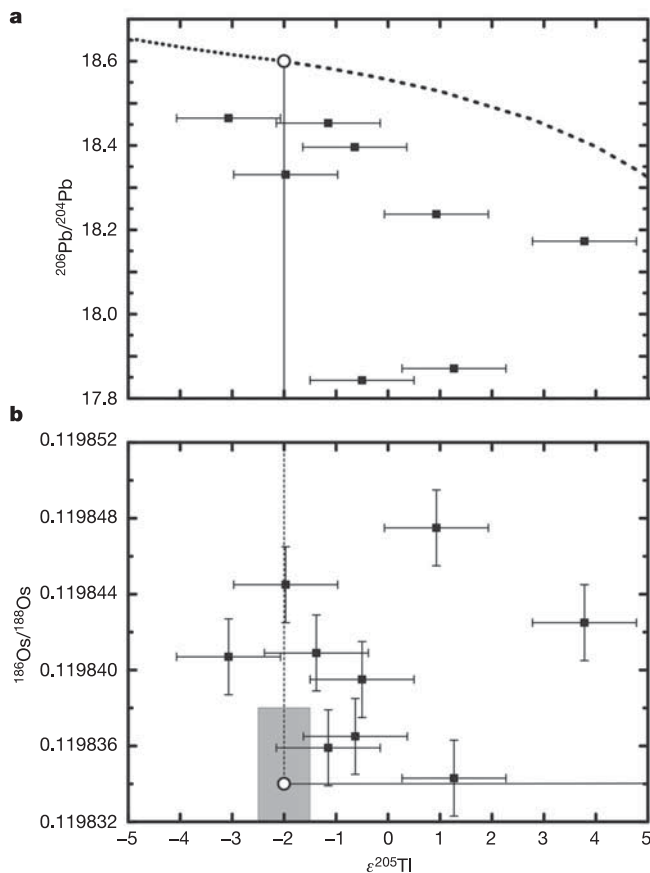


Figure 2 | Thallium isotope compositions of the Hawaiian picrites plotted versus $^{206}\text{Pb}/^{204}\text{Pb}$ and $^{186}\text{Os}/^{188}\text{Os}$ isotope ratios. Error bars denote 2 s.d. uncertainties. **a**, Lead isotope data are from refs 15, 24. Also shown are mixing lines between pristine mantle (open circle) and Fe-Mn sediments (dashed line), a lithogenic sediment component (solid line) and altered MORB (dotted line). The endmember compositions used to construct the mixing lines are as follows. Fe-Mn sediments: Pb = $500 \mu\text{g g}^{-1}$, $^{206}\text{Pb}/^{204}\text{Pb} = 16$, Tl = $100 \mu\text{g g}^{-1}$, $\varepsilon^{205}\text{Tl} = 10$; lithogenic sediment component: Pb = $30 \mu\text{g g}^{-1}$, $^{206}\text{Pb}/^{204}\text{Pb} = 17.5$, Tl = $0.6 \mu\text{g g}^{-1}$, $\varepsilon^{205}\text{Tl} = -2$; altered MORB: Pb = $0.23 \mu\text{g g}^{-1}$, $^{206}\text{Pb}/^{204}\text{Pb} = 21$, Tl = $0.05 \mu\text{g g}^{-1}$, $\varepsilon^{205}\text{Tl} = -15$; pristine mantle, Pb = $0.071 \mu\text{g g}^{-1}$, $^{206}\text{Pb}/^{204}\text{Pb} = 18.6$, Tl = $0.0012 \mu\text{g g}^{-1}$, $\varepsilon^{205}\text{Tl} = -2$. It is notable that addition of the lithogenic component has no effect on either Tl isotopes or the Cs/Tl ratio as this component is identical to the mantle values ($\varepsilon^{205}\text{Tl} \approx -2$ and Cs/Tl ≈ 6) of these two parameters. **b**, The $^{186}\text{Os}/^{188}\text{Os}$ isotope data are from refs 1, 2. Also shown are mixing lines between mantle (open circle) and Fe-Mn sediments (solid line) and core (dashed line). The grey box indicates the range of normal mantle values. The endmember compositions used to construct the mixing lines are as follows. Fe-Mn sediments: Os = $0.003 \mu\text{g g}^{-1}$, $^{186}\text{Os}/^{188}\text{Os} = 0.1206$, Tl = $100 \mu\text{g g}^{-1}$, $\varepsilon^{205}\text{Tl} = 10$; core, Os = $0.3 \mu\text{g g}^{-1}$, $^{186}\text{Os}/^{188}\text{Os} = 0.11987$, Tl = $0.005 \mu\text{g g}^{-1}$, $\varepsilon^{205}\text{Tl} = -2$; mantle, Os = $0.003 \mu\text{g g}^{-1}$, $^{186}\text{Os}/^{188}\text{Os} = 0.119834$, Tl = $0.0012 \mu\text{g g}^{-1}$, $\varepsilon^{205}\text{Tl} = -2$.

Both of these components must be characterized by unradiogenic $^{206}\text{Pb}/^{204}\text{Pb}$, which is indicative of an old (>2.5 Gyr) age and low time-integrated U/Pb ratios. These conclusions are in accord with the results of a previous study, which suggested that the mantle source of Hawaiian volcanoes was contaminated by ancient (~ 3 Gyr) pelagic sediments²⁶.

It is notable that the very small amounts of Fe-Mn sediments that are required to account for the Tl isotope variations of the Hawaiian picrites will not significantly perturb the Nd and Hf isotope systematics of the samples, and even the Pb isotope compositions are altered only marginally (Fig. 2a). Also, we note that the Fe/Mn ratio is unaffected by such small ferromanganese sediment additions. For example, admixing of $30 \mu\text{g g}^{-1}$ of Fe-Mn sediment will only change the Fe/Mn of a mantle source from 59.6 (ref. 23) to 59.4. Hence the high Fe/Mn ratios recorded by these samples¹⁰ cannot be related to addition of Fe-Mn sediments. In contrast, it was previously suggested that the $^{186}\text{Os}/^{188}\text{Os}$ isotope systematics of the Hawaiian mantle source may be dominated by admixed Fe-Mn sediments³. It is evident, however, that our data exhibit no correlation between Tl and Os isotope compositions (Fig. 2b). This observation is readily explained by the extremely high Tl/Os ratios of about 40,000 (refs 3, 11) for Fe-Mn deposits. Minor additions of such material into a normal mantle source will generate large Tl isotope anomalies, whereas $^{186}\text{Os}/^{188}\text{Os}$ ratios remain constant (Fig. 2b). If mixing processes were to be responsible for the isotope systematics of sample LO 02-02 (which displays a significant ^{186}Os isotope anomaly), this would require a contaminant characterized by Tl/Os ≈ 2 . In principle, such a contaminant could be produced from recycled Fe-Mn sediments, by preferential loss of Tl from subducting slabs. Reducing the Tl/Os ratio from 40,000 to 2 requires that more than 99.9% of the original Tl budget is lost during subduction, however. This is unrealistic, given that Tl is less fluid mobile than Pb (ref. 27), and the latter element is typically depleted by 10–90% through fluid mobilization in subduction zones²⁸. The Tl and $^{186}\text{Os}/^{188}\text{Os}$ isotope variations of the Hawaiian picrites are therefore unlikely to have the same origin, and this precludes Fe-Mn sediments as the source of the ^{186}Os isotope anomalies. This interpretation implies that core-mantle interaction is still a viable explanation for the elevated $^{186}\text{Os}/^{188}\text{Os}$ isotope ratios observed for the Hawaiian mantle plume.

METHODS

The Hawaiian picrites and Icelandic basalts were obtained as powders and analysed as received. The MORB glasses were crushed in an agate mortar to chips $<300 \mu\text{m}$ in size, which were handpicked to obtain pieces devoid of any alteration products. Following dissolution of the samples (~ 1 g), Tl was separated from the sample matrix by a two-stage column chemistry procedure previously described in refs 14 and 29. The Tl isotope analyses were performed by multiple collector inductively coupled plasma mass spectrometry at the ETH Zurich, using both external normalization to SRM 981 Pb and standard sample bracketing for mass bias correction²⁹. The external reproducibility (2 s.d.) of the Tl isotope analyses is about $\pm 1 \varepsilon^{205}\text{Tl}$ -unit^{14,29}. Procedural blanks for silicate dissolutions and column chemistry are less than 20 pg of Tl (ref. 14), which is $<1\%$ of the Tl analysed in any sample in this study and therefore insignificant. The Tl concentrations were determined by monitoring the ^{205}Tl signal intensities during the isotopic measurements and normalizing these to the ^{208}Pb ion beam from the known quantity of SRM 981 Pb added to the sample solutions. It was also taken into account that Tl is ionized approximately 5% more efficiently than Pb. These data are estimated to be accurate to better than $\pm 25\%$ (ref. 12).

The picrite samples analysed in this study were all selected carefully for their petrographic freshness, with the exception of KOO-17a, which probably has been subaerially altered¹⁵. However, subaerial weathering is not expected to cause significant Tl isotope fractionation¹⁴, though the effect on the Cs/Tl ratio is unknown. The unaltered nature of the samples is also attested by their mantle-like Rb/Cs ratios^{15,30}, which are known to change significantly during both submarine and subaerial alteration³¹.

The number of MORB analyses was limited to five (Table 1) because large samples (~ 1 g) of handpicked and extremely pure glass are necessary for the acquisition of accurate and precise Tl isotope data. The sample size is constrained by the low Tl concentrations of MORB. Only the purest MORB glass

that is completely devoid of Fe-Mn oxyhydroxide coatings or any signs of weathering (for example, devitrification) can be used for analysis, as such alteration products are known to be highly enriched in Tl (refs 12, 32) that is isotopically fractionated^{11,13} (by up to about $10\ \epsilon^{205}\text{Tl}$) relative to MORB.

The accretion model used to calculate the Tl isotope composition of the Earth's core is from ref. 21. This model features concomitant collision-driven planetary growth and core formation, where the Earth reaches 99% of its current mass at the putative giant impact that formed the Moon. The mean life of accretion, τ , was set at 15 Myr, which results in a giant impact at 43.5 Myr. The bulk Earth Pb and Tl concentrations were 2,400 and $20\ \text{ng g}^{-1}$, respectively. The model parameters used results in a silicate Earth with slightly too high Tl concentration and isotope composition. However, fitting the parameters (that is, the bulk Earth Pb and Tl concentrations) to accommodate the silicate Earth invariably leads to a core with $\epsilon^{205}\text{Tl} < 4$, even at $\tau = 5$ Myr. Thus, the calculation is made solely for illustrative purposes.

Received 4 August; accepted 18 November 2005.

- Brandon, A. D., Norman, M. D., Walker, R. J. & Morgan, J. W. ^{186}Os - ^{187}Os systematics of Hawaiian picrites. *Earth Planet. Sci. Lett.* **174**, 25–42 (1999).
- Brandon, A. D., Walker, R. J., Morgan, J. W., Norman, M. D. & Prichard, H. M. Coupled ^{186}Os and ^{187}Os evidence for core-mantle interaction. *Science* **280**, 1570–1573 (1998).
- Baker, J. A. & Jensen, K. K. Coupled ^{186}Os - ^{187}Os enrichments in the Earth's mantle - core-mantle interaction or recycling of ferromanganese crusts and nodules? *Earth Planet. Sci. Lett.* **220**, 277–286 (2004).
- Schersten, A., Elliott, T., Hawkesworth, C. & Norman, M. Tungsten isotope evidence that mantle plumes contain no contribution from the Earth's core. *Nature* **427**, 234–237 (2004).
- Ravizza, G., Blusztajn, J. & Prichard, H. M. Re-Os systematics and platinum-group element distribution in metalliferous sediments from the Troodos ophiolite. *Earth Planet. Sci. Lett.* **188**, 369–381 (2001).
- Hofmann, A. W. & White, W. M. Mantle plumes from ancient oceanic crust. *Earth Planet. Sci. Lett.* **57**, 421–436 (1982).
- Brandon, A. D. & Walker, R. J. The debate over core-mantle interaction. *Earth Planet. Sci. Lett.* **232**, 211–225 (2005).
- Walker, R. J. *et al.* Applications of the ^{190}Pt - ^{186}Os isotope system to geochemistry and cosmochemistry. *Geochim. Cosmochim. Acta* **61**, 4799–4807 (1997).
- Walker, R. J., Morgan, J. W. & Horan, M. F. Osmium-187 enrichment in some plumes: evidence for core-mantle interaction? *Science* **269**, 819–822 (1995).
- Humayun, M., Qin, L. & Norman, M. D. Geochemical evidence for excess iron in the mantle beneath Hawaii. *Science* **306**, 91–94 (2004).
- Rehkämper, M. *et al.* Thallium isotope variations in seawater and hydrogenetic, diagenetic, and hydrothermal ferromanganese deposits. *Earth Planet. Sci. Lett.* **197**, 65–81 (2002).
- Rehkämper, M., Frank, M., Hein, J. R. & Halliday, A. Cenozoic marine geochemistry of thallium deduced from isotopic studies of ferromanganese crusts and pelagic sediments. *Earth Planet. Sci. Lett.* **219**, 77–91 (2004).
- Nielsen, S. G. *et al.* An elemental and isotopic study of the marine geochemistry of thallium. *Eos* **84**, OS42F–02 (2004).
- Nielsen, S. G. *et al.* The thallium isotope composition of the upper continental crust and rivers — An investigation of the continental sources of dissolved marine thallium. *Geochim. Cosmochim. Acta* **69**, 2007–2019 (2005).
- Norman, M. D. & Garcia, M. O. Primitive magmas and source characteristics of the Hawaiian plume: petrology and geochemistry of shield picrites. *Earth Planet. Sci. Lett.* **168**, 27–44 (1999).
- Heinrichs, H., Schulzdobrick, B. & Wedepohl, K. H. Terrestrial geochemistry of Cd, Bi, Tl, Pb, Zn and Rb. *Geochim. Cosmochim. Acta* **44**, 1519–1533 (1980).
- Hein, J. R., *et al.* in *Handbook of Marine Mineral Deposits* (ed. Cronan, D. S.) 239–280 (CRC Press, Boca Raton, 2000).
- Norman, M. D., Garcia, M. O. & Bennett, V. C. Rhenium and chalcophile elements in basaltic glasses from Ko'olau and Moloka'i volcanoes: Magmatic outgassing and composition of the Hawaiian plume. *Geochim. Cosmochim. Acta* **68**, 3761–3777 (2004).
- Gauthier, P. J. & Le Cloarec, M. F. Variability of alkali and heavy metal fluxes released by Mt. Etna volcano, Sicily, between 1991 and 1995. *J. Volcanol. Geotherm. Res.* **81**, 311–326 (1998).
- Nielsen, S. G., Rehkämper, M. & Halliday, A. N. An internal ^{205}Pb - ^{205}Tl isochron for the iron meteorite Toluca and the initial Solar System abundance of ^{205}Pb . *Eos* **85**, F1251 (2004).
- Halliday, A. N. Mixing, volatile loss and compositional change during impact-driven accretion of the Earth. *Nature* **427**, 505–509 (2004).
- Wasson, J. T. & Kallemeyn, G. W. Compositions of chondrites. *Phil. Trans. R. Soc. Lond. A* **325**, 535–544 (1988).
- McDonough, W. F. & Sun, S.-s. The composition of the Earth. *Chem. Geol.* **120**, 223–253 (1995).
- Bennett, V. C., Esat, T. M. & Norman, M. D. Two mantle-plume components in Hawaiian picrites inferred from correlated Os-Pb isotopes. *Nature* **381**, 221–224 (1996).
- Lassiter, J. C. & Hauri, E. H. Osmium-isotope variations in Hawaiian lavas: evidence for recycled oceanic lithosphere in the Hawaiian plume. *Earth Planet. Sci. Lett.* **164**, 483–496 (1998).
- Blichert-Toft, I., Frey, F. A. & Albareda, F. Hf isotope evidence for pelagic sediments in the source of Hawaiian basalts. *Science* **285**, 879–882 (1999).
- Noll, P. D., Newsom, H. E., Leeman, W. P. & Ryan, J. G. The role of hydrothermal fluids in the production of subduction zone magmas: Evidence from siderophile and chalcophile trace elements and boron. *Geochim. Cosmochim. Acta* **60**, 587–611 (1996).
- Bach, W., Peucker-Ehrenbrink, B., Hart, S. R. & Blusztajn, J. S. Geochemistry of hydrothermally altered oceanic crust: DSDP/ODP Hole 504B—Implications for seawater-crust exchange budgets and Sr- and Pb-isotopic evolution of the mantle. *Geochim. Geophys. Geosyst.* **4**, 8904 (2003).
- Nielsen, S. G., Rehkämper, M., Baker, J. & Halliday, A. N. The precise and accurate determination of thallium isotope compositions and concentrations for water samples by MC-ICPMS. *Chem. Geol.* **204**, 109–124 (2004).
- Eggs, S. M. *et al.* A simple method for the precise determination of ≥ 40 trace elements in geological samples by ICPMS using enriched isotope internal standardisation. *Chem. Geol.* **134**, 311–326 (1997).
- Hofmann, A. W. & White, W. M. Ba, Rb and Cs in the Earth's mantle. *Z. Naturforsch.* **38**, 256–266 (1983).
- Jochum, K. P. & Verma, S. P. Extreme enrichment of Sb, Tl and other trace elements in altered MORB. *Chem. Geol.* **130**, 289–299 (1996).
- Ben Othmann, D., White, W. M. & Patchett, J. The geochemistry of marine sediments, island arc magma genesis, and crust-mantle recycling. *Earth Planet. Sci. Lett.* **94**, 1–21 (1989).

Acknowledgements We thank T. Elliott, M. Humayun and R.J. Walker for reviews; A. Brandon for comments; C. Devey, D. Graham, K. Haase, C. Hémond, P. Stoffers and the POS 221 and SO157 scientific parties for providing samples; M. Meier, U. Menet, D. Niederer, B. Rüttsche, C. Stirling, A. Süsli, S. Woodland, H. Williams and the rest of the IGMR group at the ETH for keeping the mass spectrometers and clean laboratories functioning at all times. This study was funded by the ETH Zurich, the Schweizerische Nationalfonds and the Danish Research Agency. This is a publication of the ARC National Key Centre for Geochemical Evolution and Metallogeny of Continents (GEMOC).

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to S.G.N. (sn Nielsen@els.mq.edu.au).

LETTERS

Tetrapod-like middle ear architecture in a Devonian fish

Martin D. Brazeau¹ & Per E. Ahlberg¹

Few fossils show the incipient stages of complex morphological transformations¹. For example, the earliest stages in the remodeling of the spiracular tract and suspensorium (jaw suspension) of osteolepiform fishes^{2–4} into the middle ear of tetrapods have remained elusive³. The most primitive known tetrapods show a middle ear architecture that is very different from osteolepiforms such as *Eusthenopteron*³, with little indication of how this transformation took place. Here we present an analysis of tetrapod middle ear origins that is based on a detailed study of *Panderichthys*, the immediate sister taxon of tetrapods. We show that the spiracular region is radically transformed from osteolepiforms and represents the earliest stages in the origin of the tetrapod

middle ear architecture. The posterior palatoquadrate of *Panderichthys* is completely tetrapod-like and defines a similarly tetrapod-like spiracular tract. The hyomandibula has lost its distal portion, representing a previously unrecognized advance towards a stapes-like morphology. This spiracular specialization suggests that the middle ear of early tetrapods evolved initially as part of a spiracular breathing apparatus^{5,6}.

The spiracular tract of fishes is bounded by the lateral wall of the braincase, the hyomandibula, and the mesial face of the palatoquadrate. The palatoquadrate governs nearly the whole lateral wall of the tract and is therefore crucial in defining the architecture of the spiracle or middle ear. In osteolepiforms (the paraphyletic group

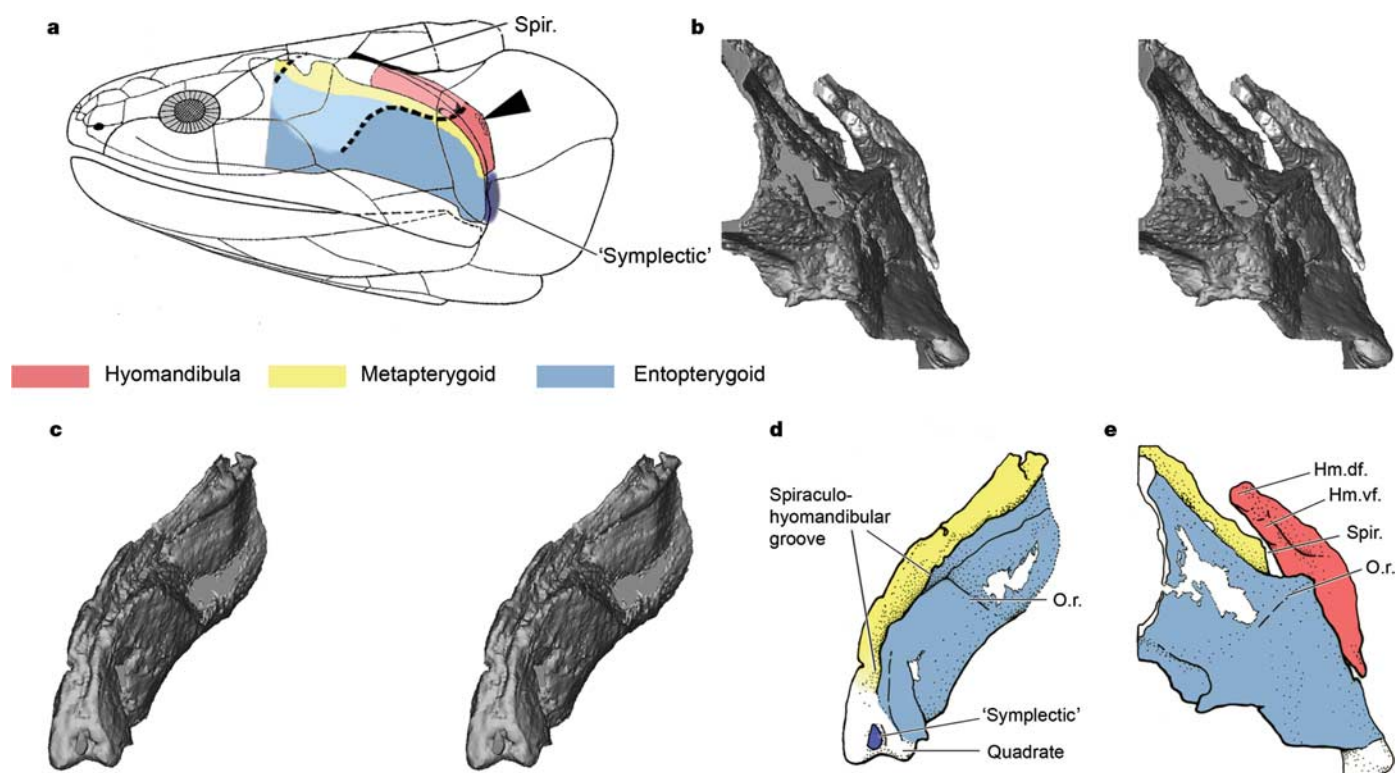


Figure 1 | Hyomandibula and subdivisions of the palatoquadrate of *Eusthenopteron foordi*. Figure is based on computed tomography scans of MHMN 06-538 showing the 'osteolepiform' condition ancestral for tetrapods. **a**, Left lateral view of skull showing, in colour, the region of the palatoquadrate recorded in the scan (modified, with permission, from ref. 26). Lighter coloured area bounded by broken lines shows extent of spiracle. **b**, Stereopair of a three-dimensional computer model of the right

palatoquadrate in oblique antero-ventro-mesial view with hyomandibula in place. **c**, Stereopair of a computer model of the left palatoquadrate in oblique postero-dorso-mesial view. **d**, **e**, Interpretive drawings of the computer models in **c** and **b**, respectively. Arrowhead indicates the position of the hyomandibula–opercula linkage. Hm.df, dorsal foot of hyomandibula; Hm.dv, ventral foot of hyomandibula; O.r., oblique ridge; Spir., spiracle.

¹Subdepartment of Evolutionary Organismal Biology, Department of Physiology and Developmental Biology, Evolutionary Biology Centre, Uppsala University, Norbyvägen 18A, SE-752 36 Uppsala, Sweden.

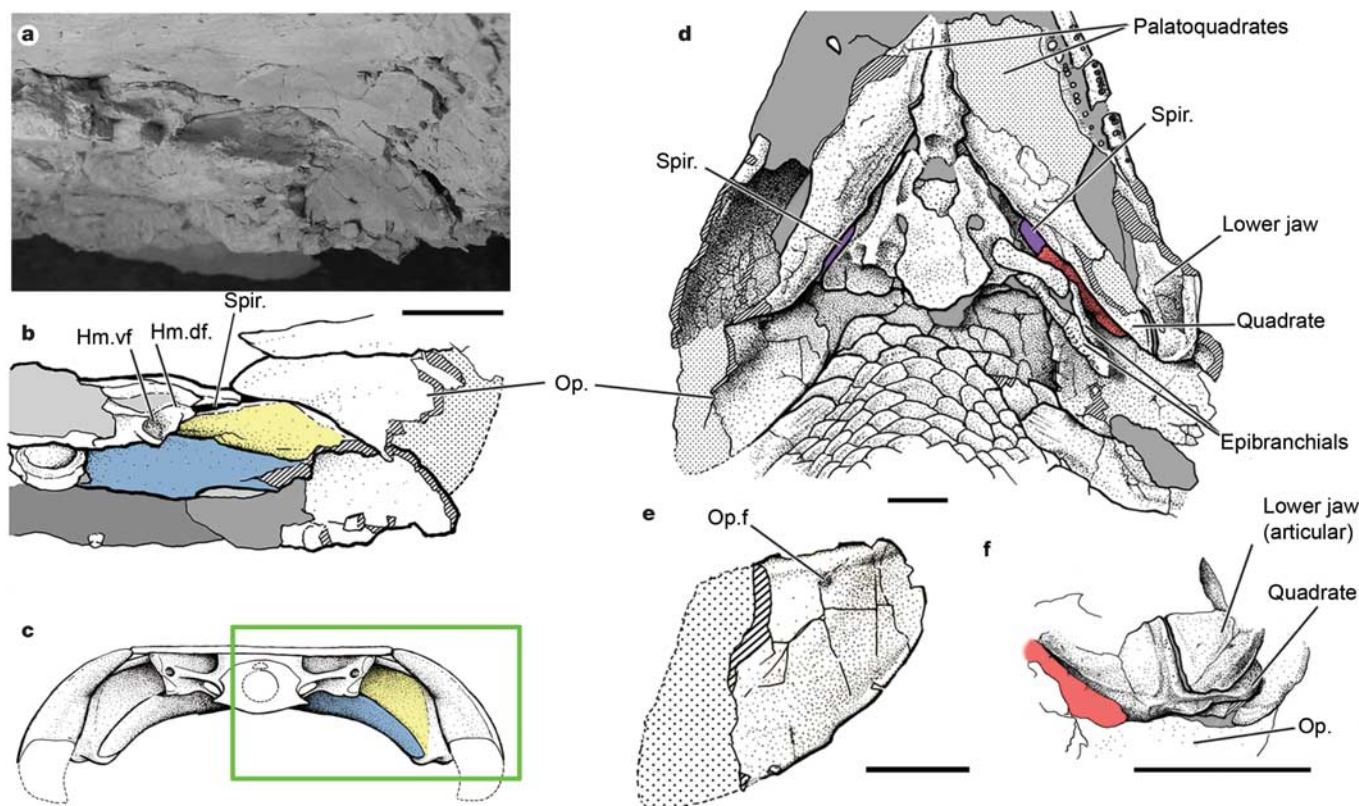


Figure 2 | Suspensorium of *Panderichthys* LDM 60/123 (Middle to Late Devonian). **a, b,** Photograph and interpretive drawing of the right palatoquadrate in posterior view. **c,** Reconstruction of skull in posterior view (with location of **a** and **b** outlined in green). **d,** Skull in palatal view showing large, straight spiracular canals. The hyomandibula (red) terminates at the level of the opercular facet. **e,** The right opercular showing opercular facet (Op.f.). **f,** Left quadrate in postero-mesial view. The hyomandibula can be

seen in close association with the opercular. Diagonal lines indicate broken bone surface; mechanical stippling indicates natural mould; grey indicates matrix (the three shades represent depth in the viewing plane, where light grey is the closest). Abbreviations and colours are the same as in Fig. 1; purple indicates spiracular opening in dermal skull roof. Scale bar, 2 cm.

ancestral to tetrapods⁷) and porolepiforms (stem-group lungfishes⁸), the palatoquadrate morphology is well known^{9–12}, giving us a clear understanding of the spiracular condition that was ancestral to the tetrapod middle ear. This condition proves to be very different to that in *Panderichthys*.

As in most basal osteichthyans¹³, the osteolepiform palatoquadrate is deep and laterally flattened posteriorly, with an entopterygoid bone forming an expansive denticle field over its mesial (buccal) surface (Fig. 1). The dorsal part of the posterior palatoquadrate, the metapterygoid region (or 'epipterygoid' in early tetrapod terminology¹⁴; Fig. 1) defines the anatomically lateral wall of the spiracular tract. In cross-section, it is narrow, dorsally concave and laterally deflected relative to the entopterygoid. An oblique ridge running across the buccal face of the entopterygoid demarcates two distinct zones: anterior to this ridge, the hyomandibula is separated from the palatoquadrate by a slot that transmitted the spiracle in life; posterior to it, the entopterygoid forms a long, intimate contact with the hyomandibula (Fig. 1). The oblique ridge evidently marks the posterior wall of the 'Eustachian tube' connecting the spiracular tract to the buccal cavity. The spiracular tract, extending between the buccal cavity and the small spiracular opening on the skull roof, could not have been more than a narrow, slot-like, dog-leg passage between the palatoquadrate, braincase and hyomandibula^{9–11}. Past analyses have emphasized similarities between the osteolepiform and tetrapod palatoquadrate, giving rise to the impression that it has not changed greatly during the emergence of tetrapods^{15,16}.

The palatoquadrate of *Panderichthys*, from the Givetian (Upper Devonian) of Latvia, was only briefly described¹⁷ without detailed comparisons to osteolepiforms or early tetrapods. Apart from its elongate external spiracular opening^{17,18}, *Panderichthys* has generally

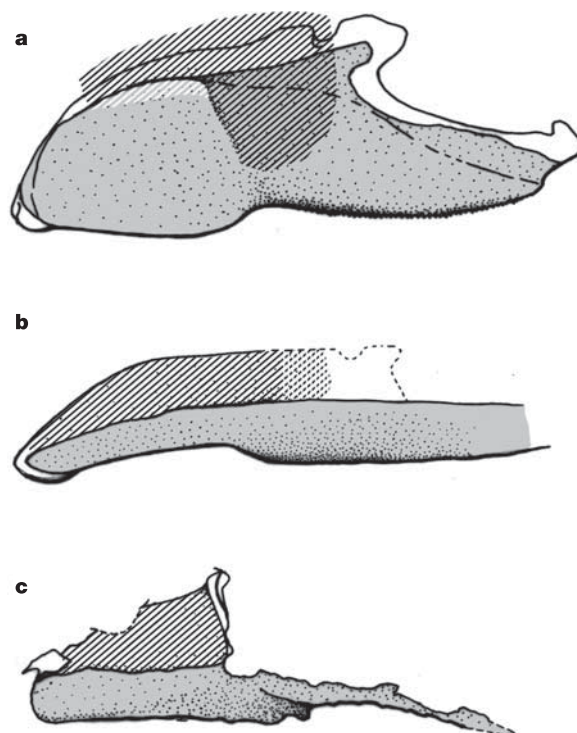


Figure 3 | Palatoquadrates. Shown are the left palatoquadrates of *Eusthenopteron* (**a**; after ref. 9), *Panderichthys* (**b**) and *Acanthostega* (**c**) in mesial view. Grey shows extent of entopterygoid; hatches show extent of the spiracular space over the palatoquadrate.

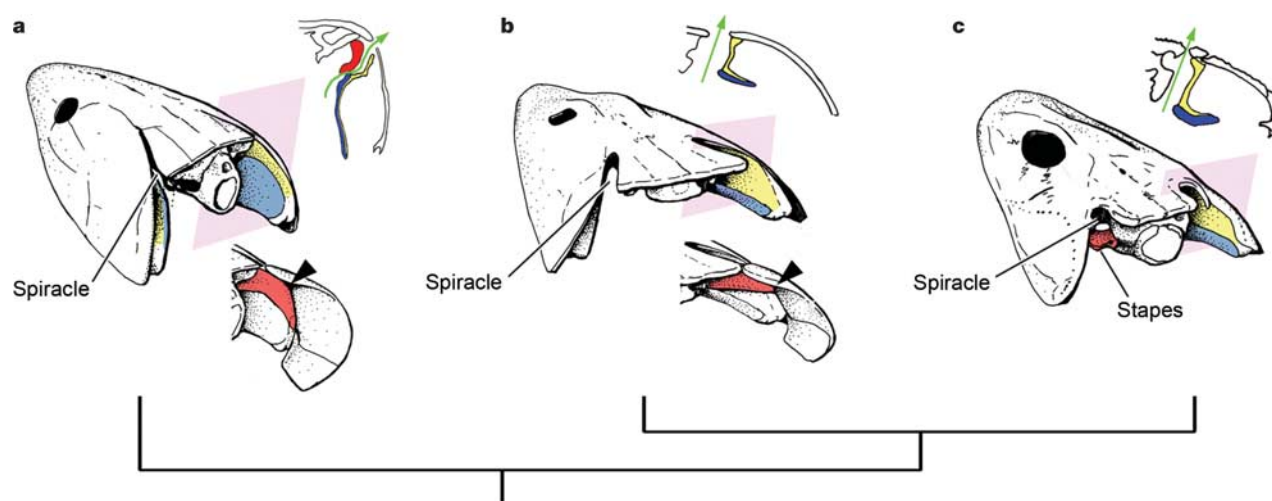


Figure 4 | Early evolution of the tetrapod middle ear space. Shown are the middle ear space of *Eusthenopteron* (a), *Panderichthys* (b) and *Acanthostega* (c) in oblique dorso-lateral view. Top images show cross-section through the spiracle at the approximate level of the pink box. Green arrow indicates the

course of the spiracle. Bottom images in a and b show suspensorium with hyomandibula and opercular series in place. Colours are the same as in Fig. 1. Not to scale.

been excluded from studies of middle ear origins because the braincase, palate and hyomandibula have been thought to be 'fish-like'^{17,18}. We examined one *Panderichthys* specimen from the Latvian Natural History Museum, Riga (LDM 60/123; Fig. 2; see Supplementary Information for additional descriptions), that is exceptional in showing these structures in ventral and posterior views¹⁸ and found that its palatoquadrate and hyomandibula in fact differ greatly from the osteolepiform condition.

In *Panderichthys* the palatoquadrate is overall much shallower, shifting the quadrate towards the level of the buccal cavity roof as in Devonian tetrapods. The posterior entopterygoid region is narrowed to a ramus that is about one-third to one-fifth of the depth of the corresponding structure in osteolepiforms (Figs 2a–c and 3). The oblique ridge and dorsal ridge of the osteolepiform entopterygoid are absent, indicating that there was no extensive zone of contact between the hyomandibula and palatoquadrate complex posterior to the spiracular tract. Coupled with the length of the external spiracular opening, this suggests that the spiracle formed an open shaft extending posteriorly almost to the hyomandibular facet on the opercular. In osteolepiforms, by contrast, the ventral and (especially) dorsal openings are smaller and more anteriorly located⁹. *Panderichthys* has a broad, flattened, nearly vertical metapterygoid region, which is much taller than the near-horizontal shelf-and-trough metapterygoid region of osteolepiforms. This creates a wider, straighter and less obstructed spiracle than in osteolepiforms. *Acanthostega*^{14,19} and *Ventastega*²⁰ (personal observation) show a palatoquadrate morphology nearly identical to that of *Panderichthys* in these respects (Figs 3 and 4).

Despite its tetrapod-like spiracular tract, *Panderichthys* lacks a true stapes^{17,18}. Like typical osteolepiforms²¹ (Fig. 1a, b, e), *Panderichthys* has a slender, rod-like hyomandibula^{17,18}. However, we observe considerable differences in the hyomandibula of *Panderichthys* (Fig. 2d, f) that show the earliest evidence of significant modification in the tetrapod stem lineage. In all other known osteichthyans with a suspensory hyomandibula, this bone⁹ (Fig. 1a) connects by means of the opercular process at mid-shaft to a facet on the mesial face of the operculum. It is evident that such a link was retained in *Panderichthys* because an opercular facet is present (Fig. 2e). However, the hyomandibula of LDM 60/123 terminates at the opercular process, rather than having the juncture at mid-shaft (Fig. 2d–f). Instead of having a long contact with the palatoquadrate complex, the hyomandibula bridges only the (posteriorly displaced) lateral commissure¹⁸ to the opercular. These modifications represent an early

step towards a free-standing, transversely oriented stapes in early tetrapods³, and suggest that the tetrapod stapes may correspond to only the proximal half of the primitive osteichthyan hyomandibula. A groove extending ventrally from the hyomandibular facet on the opercular indicates that a cartilaginous element continued the hyoid arch beyond the distal tip of the hyomandibula, perhaps forming a link to the ceratohyal (Fig. 2e).

This analysis enables us to dissect the origin of the primitive tetrapod middle ear into two discrete steps. The first step, below the *Panderichthys* node of the tetrapod stem lineage, involved transformation of the spiracular tract into wider, longer, vertically oriented space, as well as loss of the distal part of the hyomandibula and separation of this bone from the entopterygoid. The second step, between the *Panderichthys* and *Acanthostega* nodes, involved loss of the opercular bone and transformation of the hyomandibula into a transversely oriented stapes with its footplate lodged in the fenestra ovalis, but does not appear to have included any major remodelling of the spiracular tract (Fig. 4). This pattern has clear functional implications.

Because the primitive tetrapod stapedial morphology does not indicate association with a tympanum, it has been repeatedly suggested that the 'otic notch' of the earliest tetrapods in fact housed a persistent spiracle²². The data from *Panderichthys* offer strong support for this previously speculative hypothesis. As in osteolepiforms, the hyomandibula of *Panderichthys* formed a brace between the opercular and lateral commissure, and had no direct relationship to the inner ear. There is thus no reason to believe that the remodelling and enlarging of the spiracular tract below the *Panderichthys* node were an auditory adaptation. A more plausible functional model is provided by Recent benthic chondrichthyans such as rays and carpet sharks, which, like *Panderichthys*, have broad, flat skulls and bodies, subterminal mouths, dorsally facing eyes and large spiracular openings^{23,24}. These fishes use spiracular inhalation to avoid drawing grit into the gills with the respiratory current when lying on the substrate. Because the spiracular tract of *Panderichthys* seems to have formed a substantially wider and more direct conduit to the pharynx than that of osteolepiforms, and the body form implies a benthic lifestyle, we infer that the remodelling of the spiracular tract occurred to facilitate spiracular ventilation.

The apparent lack of remodelling between the spiracular spaces of *Panderichthys* and the earliest tetrapods, and the absence of a tympanum, strongly suggest that the ventilatory function of the spiracle was conserved during the origin of tetrapods. The transformation of

the spiracle into the 'middle ear' of primitive tetrapods such as *Acanthostega* involved the creation of a stapes that contacted the inner ear and may thus have acquired a rudimentary auditory role. Notably, the middle ear of *Ichthyostega*, which does appear to be adapted principally for sound transduction, is markedly divergent from the primitive tetrapod morphology²⁵. It does not seem, however, that the earliest stages in the origin of the middle ear were part of a braincase support apparatus² or involved in any auditory function. Rather, it seems that they related to the elaboration of the spiracular breathing apparatus and possibly the anchoring of a spiracle valve to the hyomandibula or stapes⁶.

Received 27 May; accepted 5 September 2005.

- Carroll, R. L. *Patterns and Processes of Vertebrate Evolution* (Cambridge Univ. Press, Cambridge, 1997).
- Carroll, R. L. in *The Terrestrial Environment and the Origin of Land Vertebrates* (ed. Panchen, A. L.) 293–317 (Academic, London, 1980).
- Clack, J. A. Patterns and processes in the early evolution of the tetrapod ear. *J. Neurobiol.* **53**, 251–264 (2002).
- Clack, J. A. Earliest known tetrapod braincase and the evolution of the stapes and fenestra ovalis. *Nature* **369**, 392–394 (1994).
- Panchen, A. L. On the amphibian *Crassigyrinus scoticus* Watson from the Carboniferous of Scotland. *Phil. Trans. R. Soc. Lond. B* **309**, 505–568 (1985).
- Clack, J. A. Discovery of the earliest-known tetrapod stapes. *Nature* **342**, 425–427 (1989).
- Ahlberg, P. E. & Johanson, Z. Osteolepiforms and the ancestry of tetrapods. *Nature* **395**, 792–794 (1998).
- Ahlberg, P. E. A re-examination of sarcopterygian interrelationships, with special reference to the Porolepiformes. *Zool. J. Linn. Soc.* **103**, 241–287 (1991).
- Jarvik, E. On the visceral skeleton in *Eusthenopteron* with a discussion of the parasphenoid and palatoquadrate in fishes. *Kungl. Svenska Vetensk. Handl.* **5**, 1–104 (1954).
- Long, J. A., Barwick, R. E. & Campbell, K. S. W. Osteology and functional morphology of the osteolepiform fish *Gogonasus andrewsae* Long, 1985, from the Upper Devonian Gogo Formation, Western Australia. *Rec. West. Aust. Mus.* **53** (suppl.), 1–89 (1997).
- Lebedev, O. A. Morphology of a new osteolepidid fish from Russia. *Bull. Mus. Natl. Hist. Nat. C* **17**, 287–341 (1995).
- Jarvik, E. Middle and Upper Devonian Porolepiformes from East Greenland with special reference to *Glyptolepis groenlandica* n. sp. *Meddelelser Grønland* **187**, 1–307 (1972).
- Gardiner, B. G. The relationships of the palaeoniscoid fishes, a review based on new specimens of *Mimia* and *Moythomasia* from the Upper Devonian of Western Australia. *Bull. Brit. Mus. (Nat. Hist.) Geol.* **37**, 173–428 (1984).
- Clack, J. A. *Acanthostega gunnari*, a Devonian tetrapod from Greenland: the snout, palate and ventral parts of the braincase, with a discussion of their significance. *Meddelelser Grønland Geosci.* **31**, 1–24 (1994).
- Vestoll, T. S. The origin of the tetrapods. *Biol. Rev.* **18**, 78–98 (1943).
- Rosen, D. E., Forey, P. L., Gardiner, B. G. & Patterson, C. Lungfishes, tetrapods, paleontology, and plesiomorphy. *Bull. Am. Mus. Nat. Hist.* **167**, 159–276 (1981).
- Vorobyeva, E. & Schultze, H.-P. in *Origins of the Higher Groups of Tetrapods: Controversy and Consensus* (eds Schultze, H.-P. & Trueb, L.) 68–109 (Cornell Univ. Press (Canstock), Ithaca, NY, 1991).
- Ahlberg, P. E., Clack, J. A. & Luksevics, E. Rapid braincase evolution between *Panderichthys* and the earliest tetrapods. *Nature* **381**, 61–64 (1996).
- Clack, J. A. The neurocranium of *Acanthostega gunnari* Jarvik and the evolution of the otic region in tetrapods. *Zool. J. Linn. Soc.* **122**, 61–97 (1998).
- Ahlberg, P. E., Luksevics, E. & Lebedev, O. The first tetrapod finds from the Devonian (Upper Famennian) of Latvia. *Phil. Trans. R. Soc. Lond. B* **343**, 303–328 (1994).
- Smithson, T. R. & Thomson, K. S. The hyomandibula of *Eusthenopteron foordi* Whiteaves (Pisces: Crossopterygii) and the early evolution of the tetrapod stapes. *Zool. J. Linn. Soc.* **74**, 93–103 (1982).
- Clack, J. A. *Gaining Ground: The Origin and Evolution of Tetrapods* (ed. Farlow, J. O.) (Indiana Univ. Press, Bloomington, IN, 2002).
- Rand, H. W. The functions of the spiracle of the skate. *Am. Nat.* **41**, 287–302 (1907).
- Summers, A. P. & Ferry-Graham, L. A. Ventilatory modes and mechanics of the hedgehog skate (*Leucoraja erinacea*): testing the continuous flow model. *J. Exp. Biol.* **204**, 1577–1587 (2001).
- Clack, J. A. et al. A uniquely specialized ear in a very early tetrapod. *Nature* **425**, 65–69 (2003).
- Jarvik, E. *Basic Structure and Evolution of Vertebrates* (Academic, London, 1980).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank E. Luksevics and I. Zupins for access to specimens; T. Rowe for computed tomography scan work; M. Arsénault and S. Desbiens for providing the *Eusthenopteron* specimen and, together with T. Rowe, for allowing us to publish these early results. M.D.B. is supported by a Fonds québécois de la recherche sur la nature et les technologies (FQRNT) bourse de maîtrise (graduate scholarship). P.E.A. is supported by a professorship and research grant from Vetenskapsrådet.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.D.B. (martin.brazeau@ebc.uu.se).

LETTERS

Reduced mixing generates oscillations and chaos in the oceanic deep chlorophyll maximum

Jef Huisman^{1*}, Nga N. Pham Thi^{2*}, David M. Karl³ & Ben Sommeijer²

Deep chlorophyll maxima (DCMs) are widespread in large parts of the world's oceans^{1–7}. These deep layers of high chlorophyll concentration reflect a compromise of phytoplankton growth exposed to two opposing resource gradients: light supplied from above and nutrients supplied from below. It is often argued that DCMs are stable features. Here we show, however, that reduced vertical mixing can generate oscillations and chaos in phytoplankton biomass and species composition of DCMs. These fluctuations are caused by a difference in the timescales of two processes: (1) rapid export of sinking plankton, withdrawing nutrients from the euphotic zone and (2) a slow upward flux of nutrients fuelling new phytoplankton production. Climate models predict that global warming will reduce vertical mixing in the oceans^{8–11}. Our model indicates that reduced mixing will generate more variability in DCMs, thereby enhancing variability in oceanic primary production and in carbon export into the ocean interior.

In oligotrophic waters, where the surface mixed layer is depleted of nutrients, subsurface maxima in chlorophyll concentration and phytoplankton biomass are often found (Fig. 1). Such deep chlorophyll maxima are permanent features in large parts of the tropical and subtropical oceans^{1–5}. Furthermore, seasonal DCMs commonly develop in temperate regions^{4,6} and even in the polar oceans⁷ when nutrients are depleted in the surface layer with the onset of the summer season. It is generally believed that DCMs are stable features, tracking seasonal changes in light and nutrient conditions. However, here we extend recent phytoplankton models^{12–16} to show that the phytoplankton populations of DCMs can show sustained fluctuations.

Consider a vertical water column. Let z indicate the depth in the water column. Let P denote the phytoplankton population density (number of cells per m^3). The population dynamics of the phytoplankton can be described by a reaction–advection–diffusion equation^{12–17}:

$$\begin{aligned} \frac{\partial P}{\partial t} &= \text{growth} - \text{loss} - \text{sinking} + \text{mixing} \\ &= \mu(N, I)P - mP - v \frac{\partial P}{\partial z} + \kappa \frac{\partial^2 P}{\partial z^2} \end{aligned} \quad (1)$$

where $\mu(N, I)$ is the specific growth rate of the phytoplankton as an increasing saturating function of nutrient availability N and light intensity I , m is the specific loss rate of the phytoplankton, v is the phytoplankton sinking velocity, and κ is the vertical turbulent

diffusivity. The nutrient dynamics in the water column can be described as^{12–14}:

$$\begin{aligned} \frac{\partial N}{\partial t} &= -\text{uptake} + \text{recycling} + \text{mixing} \\ &= -\alpha \mu(N, I)P + \varepsilon \alpha mP + \kappa \frac{\partial^2 N}{\partial z^2} \end{aligned} \quad (2)$$

where α is the nutrient content of the phytoplankton, and ε is the proportion of nutrient in dead phytoplankton that is recycled. We

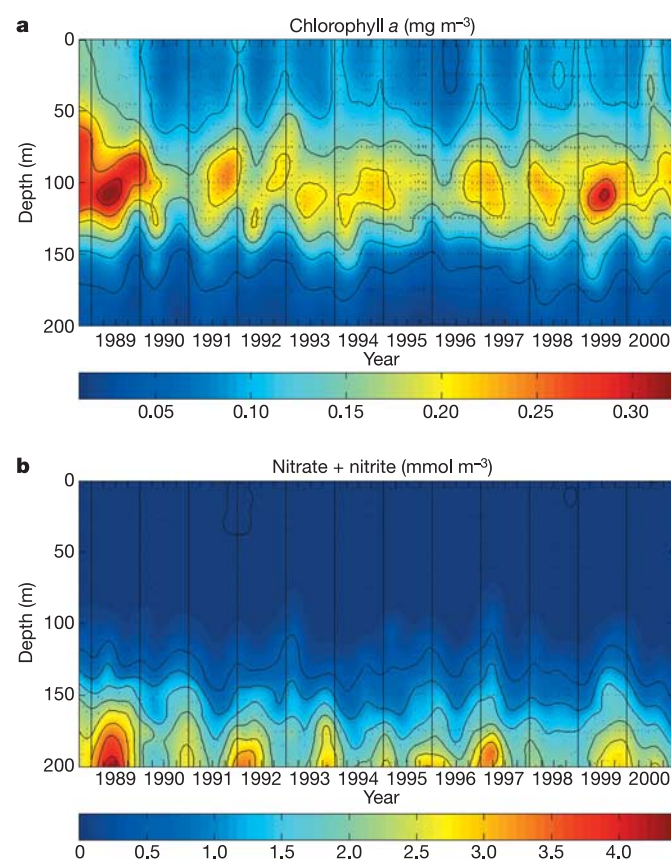


Figure 1 | Time course of the DCM at Station ALOHA, in the subtropical Pacific Ocean, North of Hawaii. **a**, Chlorophyll *a*. **b**, Nitrate and nitrite. Data were obtained from the Hawaii Ocean Time-series (HOT) program.

¹Aquatic Microbiology, Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, Nieuwe Achtergracht 127, 1018 WS Amsterdam, The Netherlands. ²Center for Mathematics and Computer Science (CWI), PO Box 94079, 1090 GB Amsterdam, The Netherlands. ³School of Ocean and Earth Science and Technology, University of Hawaii, 1000 Pope Road, Honolulu, Hawaii 96822, USA.

*These authors contributed equally to this work.

assume that light intensity, I , decreases exponentially with depth according to Lambert–Beer’s law, owing to light absorption by the phytoplankton population, by water and by dissolved substances^{15,16}. To complete the model, we use zero-flux boundary conditions for the phytoplankton. Furthermore, we assume a zero-flux boundary condition for nutrients at the surface, while nutrients are replenished from below with a fixed concentration N_B at the bottom of the water column. The model formulation and simulation methods are described in further detail in the Supplementary Information. The model is parameterized for clear ocean water, reflecting the North Pacific subtropical gyre^{5,18} (Fig. 1).

In a first model simulation, with a turbulent diffusivity of $0.5 \text{ cm}^2 \text{ s}^{-1}$, nutrients in the top layer are gradually depleted by the phytoplankton. The nutricline slowly moves downwards, tracked by the phytoplankton population, until the population settles at a stable equilibrium at which the downward flux of consumed nutrients equals the upward flux of new nutrients (Fig. 2a). Thus, a stable DCM develops. For lower values of turbulent diffusivity, however, the model predicts that the phytoplankton population in the DCM will oscillate. Depending on the parameter settings, fluctuations in the DCM can range from mild oscillations (Fig. 2b) to pronounced chlorophyll peaks (Fig. 2c). To investigate this phenomenon further, we ran numerous simulations using a wide range of turbulent diffusivities. For comparison, vertical turbulent diffusivities in the ocean interior are typically on the order of $0.1 \text{ cm}^2 \text{ s}^{-1}$ to $1 \text{ cm}^2 \text{ s}^{-1}$ (refs. 19–21). The model simulations predict that the DCM becomes unstable when turbulent diffusivity is in the lower end of the realistic range (Fig. 3a). By a cascade of period doublings, reduced turbulent mixing can even generate chaos in the DCM (Fig. 3b).

The mechanism underlying these fluctuations is a difference in timescale between the sinking flux of phytoplankton and the upward diffusive flux of nutrients. This might be called an ‘advection–

diffusion instability’. At low diffusivity, the phytoplankton sink fast compared to the slow upward flux of nutrients. Thereby, the light conditions of the sinking phytoplankton deteriorate and the phytoplankton population declines. The declining phytoplankton population loses control over the upward nutrient flux, allowing new nutrients to diffuse further upwards. The upward flux of nutrients reaches a depth at which light conditions are suitable for growth. This fuels the next peak in the DCM. Indeed, model simulations indicate that the sinking flux has an important role in these oscillations, as oscillations were not observed with neutrally buoyant phytoplankton (results not shown). The period and amplitude of the DCM oscillations increase with increasing phytoplankton sinking velocity (Fig. 3c). The period and amplitude decrease with increasing vertical diffusivity (Fig. 3d). Thus, the oscillations become more pronounced if the timescale of sinking is fast compared to the timescale of the upward flux of nutrients.

Detailed ocean time series indicate that seasonal changes in light conditions have a large effect on the dynamics of DCMs⁵ (see also Fig. 1). To add more realism to the model, we therefore forced the model by seasonal changes in incident light intensity typical for the North Pacific subtropical gyre⁵, with a winter minimum of $30 \text{ mol photons m}^{-2} \text{ d}^{-1}$ and a summer maximum of $60 \text{ mol photons m}^{-2} \text{ d}^{-1}$. At high turbulent diffusivity, the DCM tracks the seasonal changes in light conditions (Fig. 2d). When turbulent diffusivity is reduced, the DCM exhibits a phenomenon known as phase locking, in which oscillations are squeezed within the seasonal cycle (Fig. 2e). For even lower turbulent diffusivities, seasonal forcing generates irregular phytoplankton blooms with chaotic multi-annual variability (Fig. 2f). Thus, similar to findings for other nonlinear oscillators^{22,23}, fluctuating DCMs show even more complex dynamics in a seasonal environment than in a constant environment.

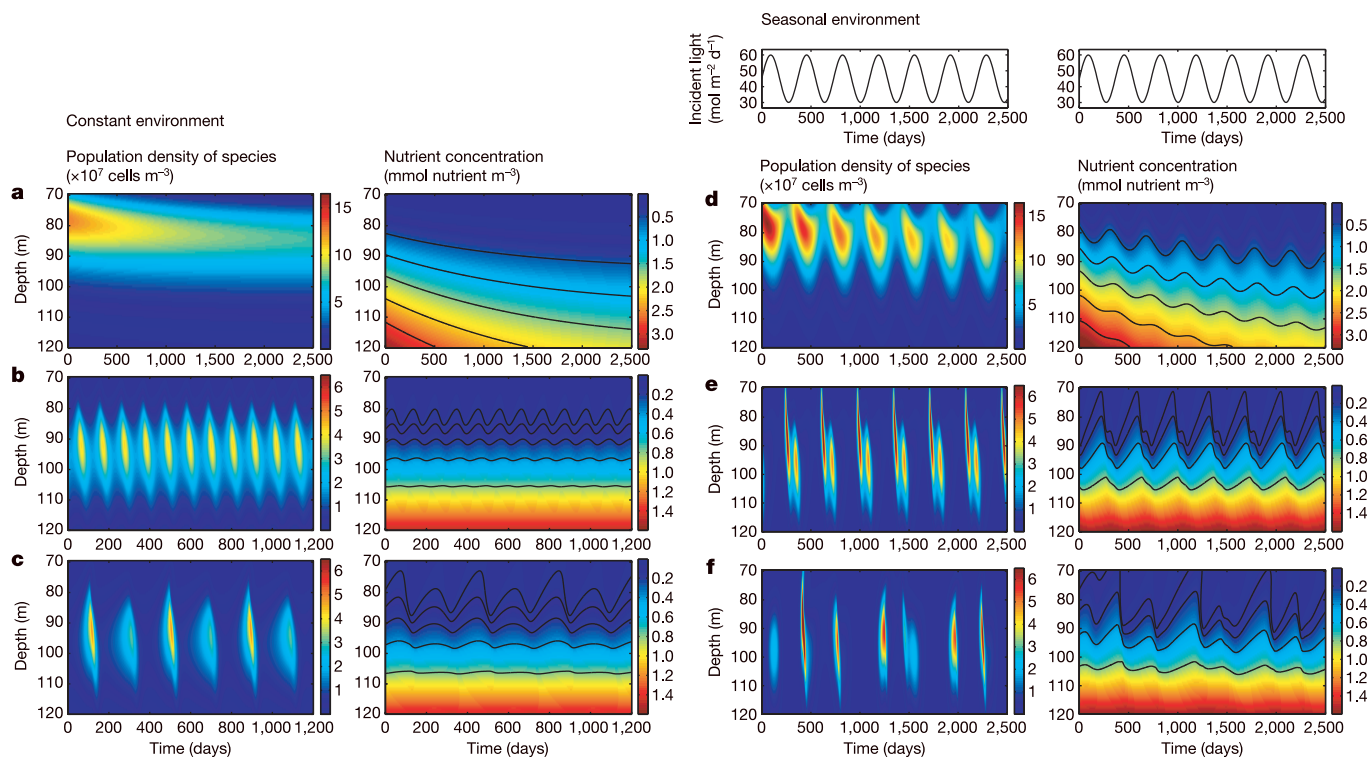


Figure 2 | Model simulations at different intensities of vertical mixing. **a–c**, Constant environment. **a**, Stable DCM ($\kappa = 0.50 \text{ cm}^2 \text{ s}^{-1}$). **b**, Mild oscillations in the DCM ($\kappa = 0.20 \text{ cm}^2 \text{ s}^{-1}$). **c**, Large-amplitude oscillations in the DCM, with double periodicity ($\kappa = 0.12 \text{ cm}^2 \text{ s}^{-1}$). **d–f**, Seasonal environment, in which the model is forced by seasonal changes in incident light intensity⁵. **d**, DCM tracks seasonal variability ($\kappa = 0.50 \text{ cm}^2 \text{ s}^{-1}$).

e, Double periodicity of DCM locked in a seasonal environment ($\kappa = 0.14 \text{ cm}^2 \text{ s}^{-1}$). **f**, Chaotic DCM in a seasonal environment ($\kappa = 0.08 \text{ cm}^2 \text{ s}^{-1}$). For **a–f**, the left panel shows phytoplankton dynamics (P) and the right panel shows nutrient dynamics (N). See Supplementary Information for parameter values.

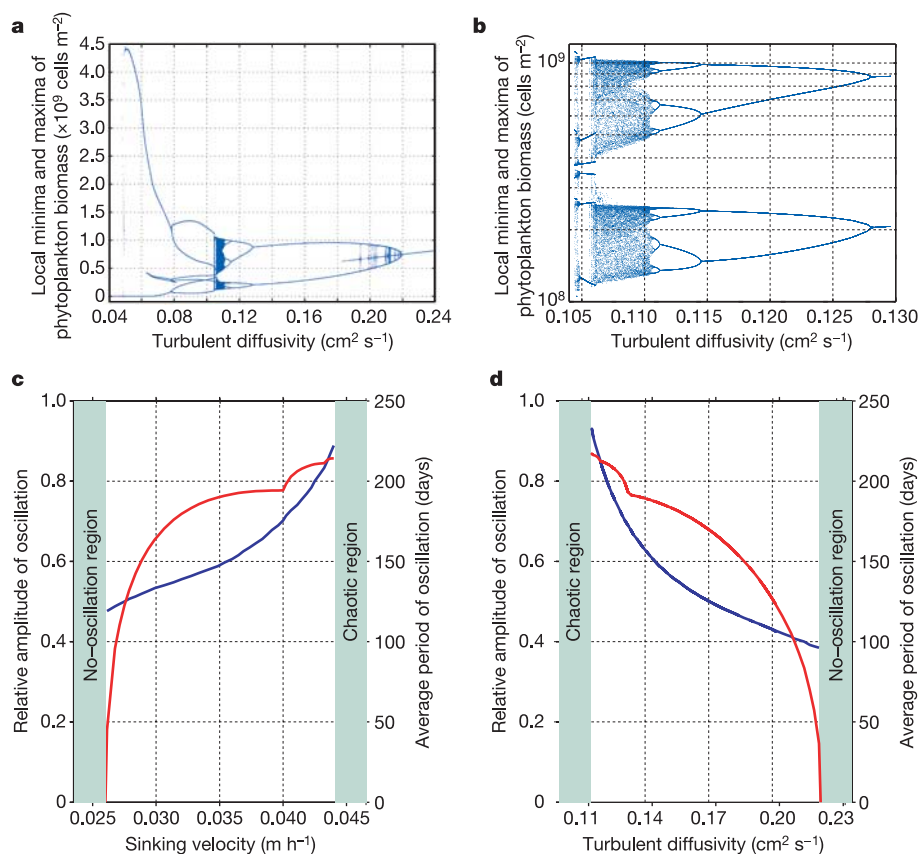


Figure 3 | Bifurcation patterns generated in a constant environment. **a**, Bifurcation diagram showing the local minima and maxima of the phytoplankton population as a function of turbulent diffusivity. **b**, Detail of the chaotic region in the bifurcation diagram. **c**, The period (blue line) and relative amplitude (red line) of the oscillations increase with phytoplankton sinking velocity. **d**, The period (blue line) and relative amplitude (red line) of the oscillations decrease with vertical turbulent diffusivity. In **a** and **b** phytoplankton population density is integrated over the upper 300 m of the water column. See Supplementary Information for parameter values.

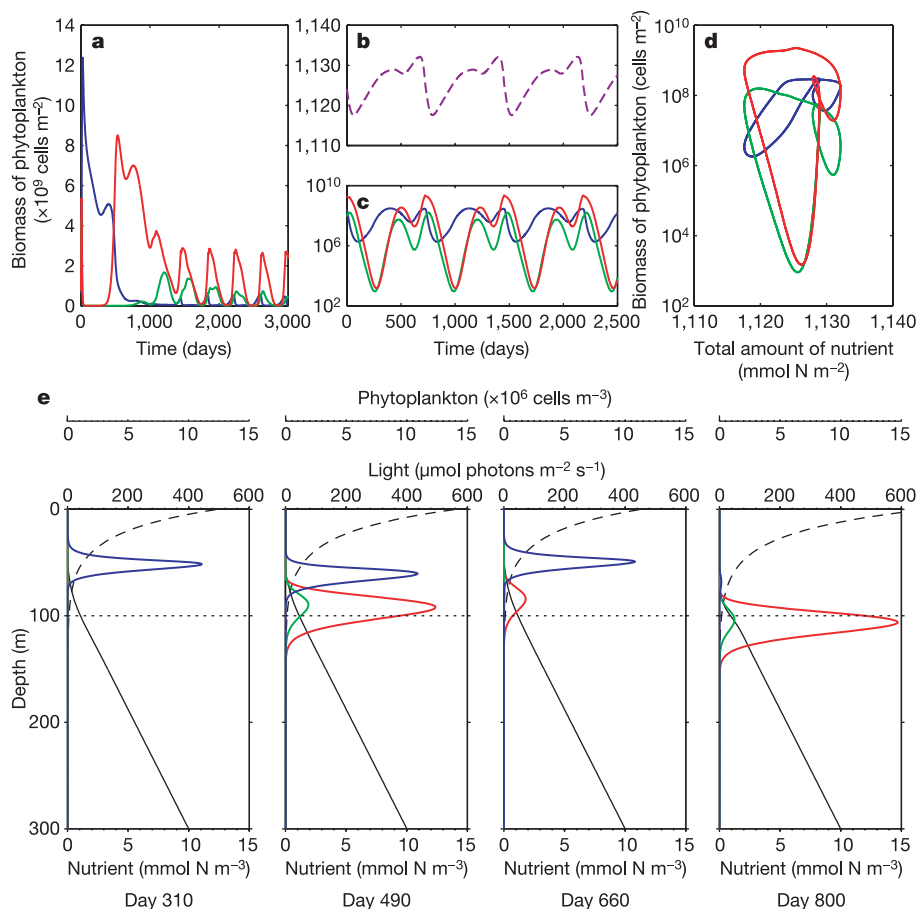


Figure 4 | Competition between three phytoplankton species in an oscillating DCM. The model (with $\kappa = 0.12 cm^2 s^{-1}$) is forced by the same seasonal changes in incident light intensity as in Fig. 2d–f. **a**, Initial time course of the phytoplankton species. **b**, **c**, In the long run, the nutrient concentration (b) and the phytoplankton species (c) settle at a periodic attractor. **d**, Phase plane illustrating the periodic attractor of the phytoplankton species. **e**, Time series of consecutive depth profiles within a single period. Coloured lines show depth profiles of the three phytoplankton species, dashed line shows light intensity, black line shows nutrient concentration. In **a**–**d** phytoplankton population density and nutrient concentration are integrated over the upper 300 m of the water column. See Supplementary Information for parameter values.

In reality, DCMs consist of multiple phytoplankton species with different growth rates, nutrient and light requirements, and sinking velocities. How would such a diverse assemblage respond to fluctuations in the DCM? To address this issue, we developed a multi-species version of our DCM model, analogous to earlier phytoplankton competition models^{16,24}. The model is again forced by seasonal changes in incident light intensity. An example is shown in Fig. 4, where we assume that the blue species has a lower sinking velocity (0.1 m d^{-1} ; resembling pico- and nanoplankton) than the red and green species (1 m d^{-1} ; resembling sinking diatoms). Furthermore, the blue species is a better nutrient competitor, whereas the red and green species are better competitors for light. Simulations show that all three species persist in this non-equilibrium environment, which confirms earlier notions that oscillations and chaos promote phytoplankton biodiversity²⁵. Periods with co-dominance of the three species are alternated with periods in which either the blue species or the red and green species dominate (Fig. 4e). Furthermore, there is a subtle but consistent vertical zonation, with the blue species (better nutrient competitor) inhabiting the nutrient-depleted upper zone of the DCM, while the red and green species (superior light competitors) peak several metres deeper in the light-deprived part of the DCM. The model predicts that phytoplankton species with relatively high sinking velocities (red and green species) show larger fluctuations than small phytoplankton species with low sinking velocities (blue species; Fig. 4c–e).

Although simple models can offer only abstractions of real-world phenomena, our model adequately reproduces many features of real-world DCMs. First, the model predicts that DCMs form at a similar depth of $\sim 100 \text{ m}$ and span a similar depth range as observed in clear ocean waters¹⁴ (Figs 1, 2). Second, consistent with observations, the model predicts that nutrients are depleted to near-zero levels above the DCM while the nutrient concentration increases linearly with depth below the DCM¹⁴ (Fig. 4e). Third, detailed ocean time-series measurements from the subtropical North Pacific confirm the prediction of a vertical zonation of species, with different species assemblages dominating at different depths²⁶ (Supplementary Information). Fourth, these ocean time series confirm the prediction that the seasonal light cycle gives rise to seasonal patterns in chlorophyll and nutrient concentrations in the DCM⁵ (Fig. 1). Fifth, the time series support the idea that plankton populations in the DCM show additional fluctuations superimposed upon the seasonal cycle, often with multi-annual variability in phytoplankton biomass and species composition^{5,18,26} (Supplementary Information). Sixth, as predicted by the model, the time series tentatively suggest that phytoplankton species with relatively high sinking velocities show larger variability than phytoplankton species with low sinking velocities (Supplementary Information). In total, time-series data support the theoretical prediction that deep chlorophyll maxima can show sustained non-equilibrium dynamics, driven by a combination of external forces and the complex internal dynamics of DCMs.

Climate models predict that global warming will increase the stability of the vertical stratification in large parts of the oceans^{8,9}. This will reduce vertical mixing and suppress the upward flux of nutrients, leading to a decline in oceanic primary production^{9–11}. Our model predicts that the same process of reduced vertical mixing may induce oscillations and chaos in the phytoplankton of the DCM, generated by the difference in timescale between the sinking flux of phytoplankton and the upward flux of nutrients. Thus, counter-intuitively, increased stability of the water column due to global warming may destabilize the phytoplankton dynamics in the DCM, with implications for oceanic primary production, species composition and carbon export.

Received 18 July; accepted 21 September 2005.

- Venrick, E. L., McGowan, J. A. & Mantyla, A. W. Deep maxima of photosynthetic chlorophyll in the Pacific Ocean. *Fishery Bull.* **71**, 41–52 (1973).
- Cullen, J. J. The deep chlorophyll maximum: comparing vertical profiles of chlorophyll *a*. *Can. J. Fish. Aquat. Sci.* **39**, 791–803 (1982).
- Mann, K. H. & Lazier, J. R. N. *Dynamics of Marine Ecosystems* (Blackwell Science, Oxford, 1996).
- Longhurst, A. R. *Ecological Geography of the Sea* (Academic, San Diego, 1998).
- Letelier, R. M., Karl, D. M., Abbott, M. R. & Bidigare, R. R. Light driven seasonal patterns of chlorophyll and nitrate in the lower euphotic zone of the North Pacific Subtropical Gyre. *Limnol. Oceanogr.* **49**, 508–519 (2004).
- Venrick, E. L. Phytoplankton seasonality in the central North Pacific: the endless summer reconsidered. *Limnol. Oceanogr.* **38**, 1135–1149 (1993).
- Holm-Hansen, O. & Hewes, C. D. Deep chlorophyll-*a* maxima (DCMs) in Antarctic waters. I. Relationships between DCMs and the physical, chemical, and optical conditions in the upper water column. *Polar Biol.* **27**, 699–710 (2004).
- Sarmiento, J. L., Hughes, T. M. C., Stouffer, R. J. & Manabe, S. Simulated response of the ocean carbon cycle to anthropogenic climate warming. *Nature* **393**, 245–249 (1998).
- Bopp, L. et al. Potential impact of climate change on marine export production. *Glob. Biogeochem. Cycles* **15**, 81–99 (2001).
- Sarmiento, J. L. et al. Response of ocean ecosystems to climate warming. *Glob. Biogeochem. Cycles* **18**, doi:10.1029/2003GB002134 (2004).
- Schmittner, A. Decline of the marine ecosystem caused by a reduction in the Atlantic overturning circulation. *Nature* **434**, 628–633 (2005).
- Fennel, K. & Boss, E. Subsurface maxima of phytoplankton and chlorophyll: steady-state solutions from a simple model. *Limnol. Oceanogr.* **48**, 1521–1534 (2003).
- Hodges, B. A. & Rudnick, D. L. Simple models of steady deep maxima in chlorophyll and biomass. *Deep-Sea Res.* **51**, 999–1015 (2004).
- Klausmeier, C. A. & Litchman, E. Algal games: the vertical distribution of phytoplankton in poorly mixed water columns. *Limnol. Oceanogr.* **46**, 1998–2007 (2001).
- Huisman, J., Arrayás, M., Ebert, U. & Sommeijer, B. How do sinking phytoplankton species manage to persist? *Am. Nat.* **159**, 245–254 (2002).
- Huisman, J. et al. Changes in turbulent mixing shift competition for light between phytoplankton species. *Ecology* **85**, 2960–2970 (2004).
- Okubo, A. & Levin, S. A. *Diffusion and Ecological Problems: Modern Perspectives* 2nd edn (Springer, Berlin, 2001).
- Karl, D. M. et al. Seasonal and interannual variability in primary production and particle flux at Station ALOHA. *Deep-Sea Res.* **43**, 539–568 (1996).
- Lewis, M. R., Harrison, W. G., Oakey, N. S., Hebert, D. & Platt, T. Vertical nitrate fluxes in the oligotrophic ocean. *Science* **234**, 870–873 (1986).
- Smyth, W. D., Moum, J. N. & Caldwell, D. R. The efficiency of mixing in turbulent patches: inferences from direct simulations and microstructure observations. *J. Phys. Oceanogr.* **31**, 1969–1992 (2001).
- Finnigan, T. D., Luther, D. S. & Lukas, R. Observations of enhanced diapycnal mixing near the Hawaiian ridge. *J. Phys. Oceanogr.* **32**, 2988–3002 (2002).
- Rinaldi, S., Muratori, S. & Kuznetsov, Y. Multiple attractors, catastrophes and chaos in seasonally perturbed predator-prey communities. *Bull. Math. Biol.* **55**, 15–35 (1993).
- Vandermeer, J., Stone, L. & Blasius, B. Categories of chaos and fractal basin boundaries in forced predator-prey models. *Chaos Soliton Fract.* **12**, 265–276 (2001).
- Huisman, J. & Sommeijer, B. Population dynamics of sinking phytoplankton in light-limited environments: simulation techniques and critical parameters. *J. Sea Res.* **48**, 83–96 (2002).
- Huisman, J. & Weissing, F. J. Biodiversity of plankton by species oscillations and chaos. *Nature* **402**, 407–410 (1999).
- Venrick, E. L. Phytoplankton species structure in the central North Pacific, 1973–1996: variability and persistence. *J. Plankton Res.* **21**, 1029–1042 (1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank R. R. Bidigare for HPLC pigment analyses, and M. Stomp, J. G. Verwer and J. Williams for discussions. J.H. was supported by the Earth and Life Sciences Foundation (ALW), which is subsidized by the Netherlands Organization for Scientific Research (NWO). N.N.P.T. was supported by the Computational Science program of NWO. D.M.K. acknowledges support from the US National Science Foundation and the Gordon and Betty Moore Foundation. B.S. acknowledges support from the Dutch BSIK/BRICKS project.

Author Contributions J.H. and N.N.P.T. contributed equally to this work. J.H., N.N.P.T. and B.S. developed the model structure. N.N.P.T. and B.S. wrote the numerical code. D.M.K. provided data from the Hawaii Ocean Time-series program. J.H. wrote the paper. All authors discussed the results and commented on the manuscript.

Author Information The time-series data from the Hawaii Ocean Time-series program are deposited at <http://hahana.soest.hawaii.edu/hot/hot-dogs>. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.H. (jef.huisman@science.uva.nl).

LETTERS

Dissection of epistasis in oligogenic Bardet-Biedl syndrome

Jose L. Badano¹, Carmen C. Leitch¹, Stephen J. Ansley¹, Helen May-Simera⁵, Shaneka Lawson¹, Richard Alan Lewis⁶, Philip L. Beales⁵, Harry C. Dietz^{1,2}, Shannon Fisher^{1,3} & Nicholas Katsanis^{1,4}

Epistatic interactions have an important role in phenotypic variability, yet the genetic dissection of such phenomena remains challenging¹. Here we report the identification of a novel locus, *MGC1203*, that contributes epistatic alleles to Bardet-Biedl syndrome (BBS), a pleiotropic, oligogenic disorder^{2–9}. *MGC1203* encodes a pericentriolar protein that interacts and colocalizes with the BBS proteins. Sequencing of two independent BBS cohorts revealed a significant enrichment of a heterozygous C430T mutation in patients, and a transmission disequilibrium test (TDT) showed strong over-transmission of this variant. Further analyses showed that the 430T allele enhances the use of a cryptic splice acceptor site, causing the introduction of a premature termination codon (PTC) and the reduction of steady-state *MGC1203* messenger RNA levels. Finally, recapitulation of the human genotypes in zebrafish shows that modest suppression of *mgc1203* exerts an epistatic effect on the developmental phenotype of BBS morphants. Our data demonstrate how the combined use of biochemical, genetic and *in vivo* tools can facilitate the dissection of epistatic phenomena, and enhance our appreciation of the genetic basis of phenotypic variability.

Despite major progress in elucidating the genetic basis of inherited disorders, much of the observed phenotypic variability cannot be explained by mutations at a single locus, leading to the exploration of oligogenic models of disease transmission, in which multiple loci exert a synergistic effect to modify the penetrance and/or expressivity of disease traits¹.

BBS is inherited primarily as an autosomal recessive trait. However, in some patients, three mutations across two BBS loci interact to modify the onset and/or severity of the phenotype^{2–9}. To date, eight BBS genes (*BBS1*–*BBS8*) have been identified^{2,5,9–16}, and a combination of *in vivo* and *in vitro* evidence suggests that BBS is a disorder of basal bodies and cilia^{10,17}.

The degree of clinical variability in BBS is not fully reconciled by interactions between the known BBS genes¹⁸. We reasoned that loci encoding proteins pertinent to the BBS functional circuit would be strong candidates to contribute modifying alleles. To identify such loci, we performed multiple rounds of yeast two-hybrid screens and identified >60 putative interactors. However, comparison of these with our recently described ciliary proteome⁹ revealed that only a single sequence, *MGC1203* (also known as *CCDC28B*; GenBank accession number NM_024296), was present in both data sets. This computationally predicted polypeptide is composed of 241 amino acids, bears no recognizable motifs, and interacts with *BBS4* in yeast (Supplementary Fig. 1a).

To investigate this interaction, we expressed epitope-tagged *MGC1203* and *BBS4* in mammalian cells. Immunoprecipitations

with *BBS4* followed by immunoblotting for *MGC1203* yielded a single band of the predicted size of 35 kDa (Fig. 1a). This interaction was not restricted to *BBS4* but was seen for every BBS protein tested (Fig. 1a).

Next, we raised a polyclonal antibody against *MGC1203* (Supplementary Fig. 1b) and localized *MGC1203* near centrosomes and basal bodies of HeLa or IMCD3 cells, a localization pattern identical to *BBS4*, *BBS6* and *BBS8*^{10,19,20} (Fig. 1b). We also found *MGC1203* to be spatially coincident with the BBS proteins in tissues pertinent to the disorder, including retina, pericardium and limb epithelium (Fig. 1c; Supplementary Fig. 1c).

Our data suggested that *MGC1203* might be relevant to the genetic aetiology of BBS. To test this hypothesis, we confirmed its genomic structure (Supplementary Fig. 2) and screened 226 unrelated BBS patients without preselecting for mutational load in the known BBS loci. In an initial BBS cohort and ethnically matched controls, we found a C → T transversion at the penultimate position of exon 3 (C430T) present in the heterozygous state in 3/64 unrelated BBS patients, compared to 4/274 controls. Analysis of a second cohort showed an even greater enrichment, with 11/162 unrelated patients carrying the 430T allele. Overall, the 430T variant of *MGC1203* was present in 6.2% of BBS patients compared to 1.4% of controls, showing significant association with BBS (Fisher's exact test $P < 0.006$). As a second, independent, test, we screened *MGC1203* in all available parents of the BBS patients analysed. We identified 27 trios with a 430T heterozygous parent, and performed TDT analysis. We found the 430T allele transmitted to patients in 20/27 trios, deviating significantly from the expected 50:50 distribution ($P < 0.007$).

The *MGC1203* mutations are probably insufficient to cause BBS. Not only did we find no patients with homozygous or compound heterozygous *MGC1203* mutations, but we found one unaffected 430T homozygous parent. Moreover, five patients carried two mutations at a known BBS locus (Supplementary Table 1), suggesting that the observed association between the 430T allele and BBS might reflect an epistatic relationship. In three of the 14 families with the 430T allele, some, but not all, affected individuals inherited the *MGC1203* mutation (Supplementary Table 1). In each case, the unbiased clinical view (blinded to the genotype) was that 430T-bearing individuals were more severely affected. In family AR46, individual –04 (*BBS1*: Y113X/M390R; *MGC1203*: 430C/C) was first diagnosed with retinitis pigmentosa (RP) at the age of 12, whereas sibling –05 (*BBS1*: Y113X/M390R; *MGC1203*: 430C/T) presented with aggressive RP at age five, ataxia and gastroschisis. In family AR709, patient –03 (*BBS1*: M390R/E549X; *MGC1203*: 430C/C) developed RP at 12 years of age, whereas –04 (*BBS1*: M390R/E549X; *MGC1203*: 430C/T) was diagnosed with RP at six years of

¹McKusick-Nathans Institute of Genetic Medicine, ²Howard Hughes Medical Institute, ³Department of Cell Biology, ⁴Wilmer Eye Institute, Johns Hopkins University, Baltimore, Maryland 21205, USA. ⁵Molecular Medicine Unit, Institute of Child Health, University College London, London WC1N 1EH, UK. ⁶Departments of Molecular and Human Genetics, Ophthalmology, Pediatrics, and Medicine, Baylor College of Medicine, Houston, Texas 77030, USA.

age, had 20/400 vision and underwent surgery at infancy for Hirschsprung disease. Finally, in family AR151, although both patients were diagnosed with RP at similar ages (six and nine), individual -04 (*MGC1203*: 430C/T) presented with a more aggressive form of the disease with macular involvement, as well as asthma, profound speech delay, and thoracolumbar scoliosis. A fourth sibship, PB029, provided a more definitive phenotypic stratification. In

this family, both affected individuals are homozygous for the common M390R mutation in *BBS1* and are C430T heterozygotes. However, the unaffected father is also M390R homozygous, and this family has been put forth as a possible example of complex inheritance⁴ (Fig. 2a). The patients inherited the *MGC1203* 430T allele from their mother, whereas the father is homozygous 430C (Fig. 2b).

The 430T allele might be causally related to BBS or be in linkage disequilibrium with another mutation. Although this variant results in a silent change, its position in the penultimate base of exon 3 raised the possibility that it might affect splicing. To investigate this, we suppressed nonsense-mediated decay (NMD) with emetine in cell lines from two unrelated 430T heterozygous patients and sequenced *MGC1203*. Although we did not detect aberrant splicing at the 3' junction of exon 3, we found that, in addition to all normal splice isoforms of *MGC1203*, 10% of the sequenced products contained a 5-bp deletion in the acceptor site at the 5' junction of the same exon. This deletion was generated by the inappropriate utilization of a cryptic splice acceptor site (Fig. 2c) and resulted in a PTC. Analysis with ESEfinder²¹ predicted that the 430T variant might improve an exonic splice enhancer (ESE) motif recognized by the SR protein SC35 (for splicing component of 35 kDa)²² (Fig. 2c). Such proteins strengthen the binding of the spliceosome to suboptimal, upstream acceptor binding sites. To test this possibility, we performed real-time polymerase chain reaction with reverse transcription (RT-PCR) on cell lines with either the 430C/C or 430C/T genotype, with or without emetine. We found that the steady-state abundance of *MGC1203* message was greater in all emetine-treated cells, suggesting that NMD occurs irrespective of genotype. However, we found a reduction of ~20% in *MGC1203* mRNA levels in untreated 430C/T cells (Fig. 2d), suggesting that a larger fraction of mRNA from the 430C/T genotype contains a PTC.

We next constructed a series of *MGC1203* minigenes (spanning from exon 2 to exon 4) and monitored the relative transcription of the mis-spliced message (*del*) relative to total *MGC1203* transcription. The 430T allele showed a marked increase in *del* mRNA production compared to a 430C minigene from both a normal individual and a 430C/T heterozygous patient (Fig. 2e). Notably, reversion of the 430T residue to 430C restored the relative production of the *del* allele to wild-type ratios (Fig. 2e). Finally, we suppressed SC35 and ASF/SF2 (alternate splicing factor/splicing factor 2) mRNA with previously reported siRNA oligomers^{23,24} and analysed the absolute amounts of *del MGC1203* transcript. In agreement with our earlier data, we found minimal production of the mis-spliced transcript from either the 430C haplotype or the mutagenized 430T → C haplotype (2 and 8 pg, respectively) compared with 170 pg produced from the 430T minigene (Fig. 2f). Suppression of approximately 70% of SC35 or >90% of ASF/SF2 message resulted in a dramatic reduction of the deleted *MGC1203* species to 70 pg and 3 pg, respectively, suggesting that the 430T allele enhances the effect of SR proteins to potentiate the cryptic splice site (Fig. 2f).

Cumulatively, our data suggest that a hypomorphic *MGC1203* mutation in humans can exert an epistatic effect on BBS mutations. To investigate this possibility *in vivo*, we undertook an antisense morpholino strategy in zebrafish. We first established that each transcript is expressed both maternally and zygotically, and is found along the rostrocaudal axis of the embryo (Supplementary Fig. 3). We designed morpholinos against the single *BBS4* and *BBS6* orthologues, the two most extensively characterized BBS genes^{19,20}, and the single *MGC1203* orthologue. Although we did not detect appreciable numbers of late embryonic phenotypes, such as renal or heart positioning/looping defects, *bbs4* morphants showed dosage-dependent phenotypes during somitogenesis (Fig. 3; Supplementary Fig. 4). Severely affected embryos showed dorsal thinning and shorter body length. By the 12-somite stage, the notochord was kinked and sometimes twisted, accompanied by widening of the somites. Finally, we observed cell detachment along the neural tube,

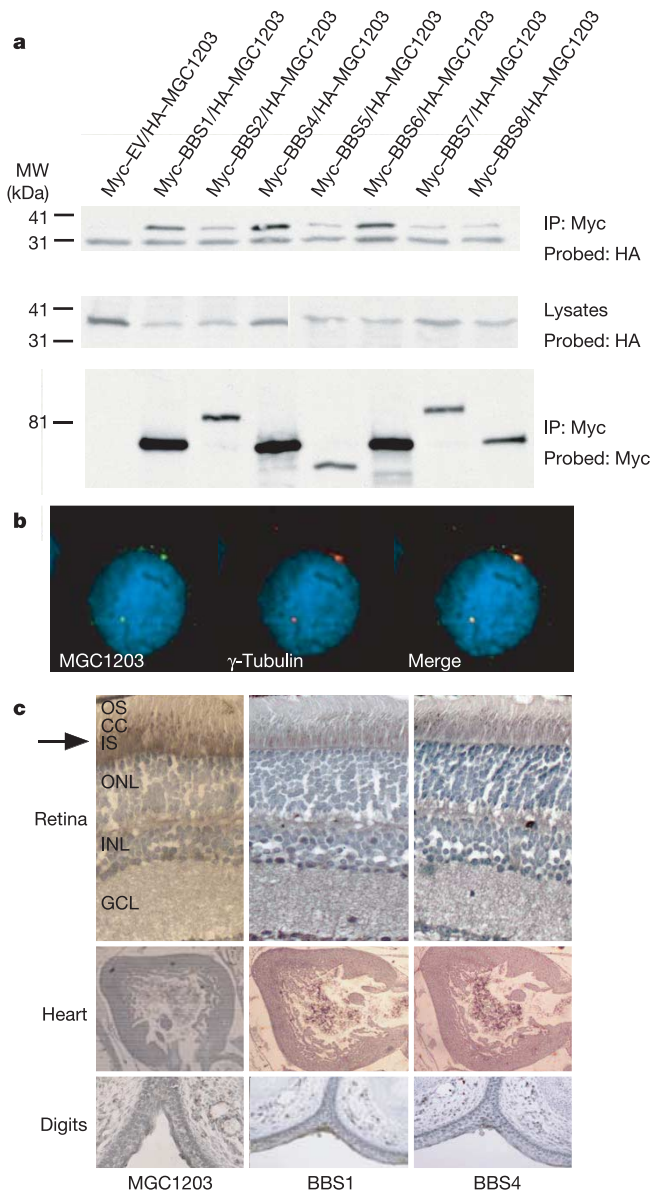


Figure 1 | *MGC1203* interacts and colocalizes with BBS proteins. **a**, HA-*MGC1203* immunoprecipitates with all BBS proteins tested (top panel). Cell lysates probed with the anti-HA antibody (middle panel) and immunoprecipitates probed with an anti-Myc monoclonal antibody (bottom panel) are shown as controls. MW, molecular weight; IP, immunoprecipitation with the indicated antibody; Probed, western blot with the indicated antibody. **b**, In IMCD3 cells, *MGC1203* (green) colocalizes with γ -tubulin (red). DNA was stained with DAPI (4,6-diamidino-2-phenylindole). **c**, *MGC1203* colocalizes with BBS1 and BBS4 in tissues. In the adult mouse retina, all proteins show overlapping patterns of expression (arrow). Similarly, all proteins are expressed ubiquitously in the pericardium of the developing heart and the epithelium surrounding the digits of embryonic day (E) 15.5 embryos. OS, outer segment; CC, connecting cilium; IS, inner segment; ONL, outer nuclear layer; INL, inner nuclear layer; GCL, ganglion cell layer.

most frequently in the head and tail regions—reminiscent of defects in neurulation and neural cell adhesion²⁵ (Fig. 3) and consistent with the recently demonstrated genetic interaction of some basal body proteins with the planar cell polarity pathway^{26,27}. These phenotypes could be rescued by co-injection of 50 pg of *bbs4* RNA, which by itself did not show any remarkable effect. We observed similar phenotypes with a *bbs6* translational-blocking morpholino, although the severity and incidence of the phenotypes was not as prominent (Supplementary Fig. 4, 5). We then injected a splice-blocking morpholino against *mgc1203* and observed phenotypes that overlapped the *bbs4* and *bbs6* defects. *Mgc1203* morphant embryos had shorter body axes but did not display appreciable dorsal thinning. They also showed moderately widened, but noticeably kinked, notochords. Somites were not widened, but were shaped abnormally and lacked definition. Finally, we observed cell detachment along the neural tube, most prominently in the head region (Fig. 3; Supplementary Fig. 4). These phenotypes correlated with the relative potency of each morpholino; mild morphants showed ~60% reduction of *mgc1203* message, whereas severe morphants had >90% reduction (Supplementary Fig. 6). These phenotypes were also specific to the morpholino as they could be rescued by co-injection of 75 pg of *mgc1203* RNA, which by itself does not produce any noticeable effects. In addition, injection of a second *mgc1203* morpholino that binds to the 5' untranslated region of *mgc1203* produced comparable phenotypes (data not shown).

We next performed double morpholino injections. Injection of either 5 ng of *mgc1203* morpholino or 3 ng of *bbs4* morpholino consistently produced >80% phenotypically normal embryos; when injected together, however, 68% of embryos had either the moderate or severe phenotypic features of a higher dose of *bbs4* morpholino alone (Fig. 3; Supplementary Fig. 7a). Notably, co-injection of 75 pg of *mgc1203* RNA with the two morpholinos rescued the double-morphant phenotype, further suggesting that the synthetic

phenotype was due to the interaction between the morpholinos, as opposed to potentially non-specific morpholino effects.

The interaction between *MGC1203* and *BBS6* was more pronounced. A subeffective dose of *mgc1203* morpholino enhanced the phenotypes caused by low doses of *bbs6* morpholino (Supplementary Fig. 7b), producing a higher incidence of BBS defects in conjunction with a low dose of *bbs6* morpholino than seen in any experiment with a higher dose of *bbs6* morpholino alone. Significantly, partial depletion of both *mgc1203* and either *bbs4* or *bbs6* produced phenotypes not found in any *mgc1203* single morphants, such as dorsal thinning and broadening of the somites.

Finally, we tested for interaction of *mgc1203* with *bbs1*. The *bbs1* morphant phenotypes produced with two non-overlapping translational-blocking morpholinos were indistinguishable from *bbs4* or *bbs6* morphants, although the number of affected embryos was modest (Fig. 4a, b). However, double *bbs1*;*mgc1203* morphants showed phenotypes that were not only reminiscent of *bbs4*;*mgc1203* and *bbs6*;*mgc1203* double injections, but were the most severe with respect to the disruption of somitic definition (Fig. 4a, b).

Our studies indicate that *MGC1203*, a protein identified through a double filter of yeast two-hybrid and computational determination of ciliary proteins, is involved in the pathogenesis of BBS by contributing hypomorphic mutations to an already sensitized genetic background. The 430T mutation exerts its effect at the RNA level by enhancing the use of a cryptic splice junction in a suboptimal context that is probably stabilized by SR proteins such as SC35 and ASF/SF2. Although there are examples in the literature of ESE mutations, the effect of such changes is typically the inefficient inclusion/exclusion of an alternatively spliced exon²¹, whereas here this phenomenon leads to the partial loss of *MGC1203* message. This model is reminiscent of erythropoietic protoporphyria, where penetrance is enhanced by the loss of a modest amount of ferrochelatase (*FECH*) mRNA owing to aberrant splicing and NMD²⁸, although in

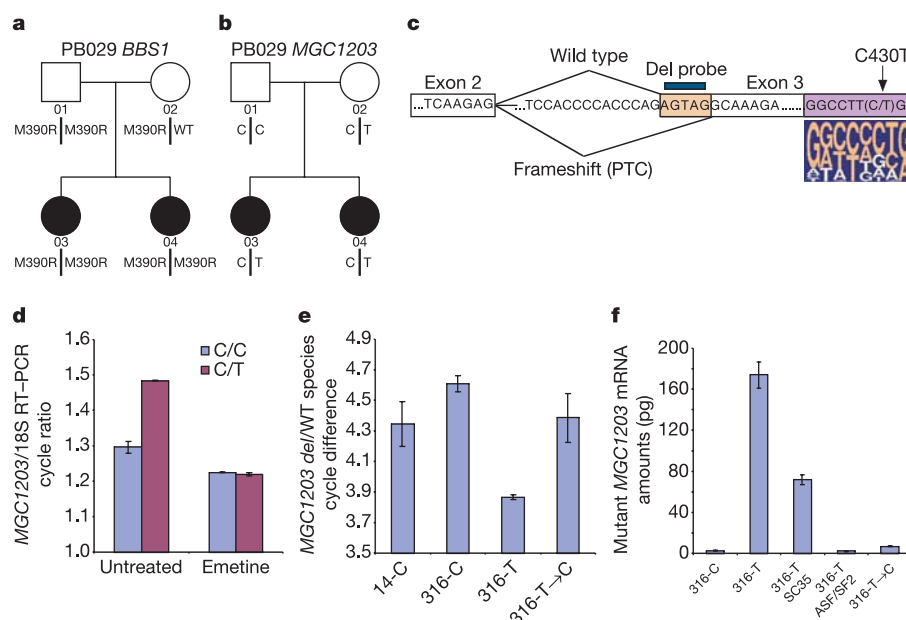


Figure 2 | Epistasis and splice defects of the 430T variant. **a, b**, In pedigree PB029, the presence of the *MGC1203* 430T allele segregates with a penetrant phenotype, as the unaffected father is homozygous mutant for *BBS1* (**a**) but wild type (WT) for *MGC1203* (**b**). **c**, In addition to wild-type *MGC1203*, a mutant transcript is also produced, containing a 5-bp deletion (orange box). The position of the C430T change is shown in the context of a putative binding site for SC35 (purple box; the consensus sequence matrix is shown underneath); the real-time PCR probe that detects the RNA deletion is indicated (Del probe). **d**, Real-time RT-PCR results with a wild-type (WT) probe with or without emetine in lymphoblastoid cells from 430C/C and

430C/T individuals. The cycles needed to reach a determined threshold are used to measure template abundance, normalized to 18S RNA. **e**, The abundance of *MGC1203* del/WT mRNA species is measured by the difference in PCR cycles between the two assays. Mutagenizing the T allele base back to a C restores the del/WT balance observed in the 430C minigenes from both backgrounds (14-C and 316-C). **f**, Quantification of del amounts using a plasmid standard curve yielded comparable results; suppression of SC35 or ASF/SF2 significantly reduces the amount of del mRNA. Experiments were performed in triplicate; Error bars are s.e.m.

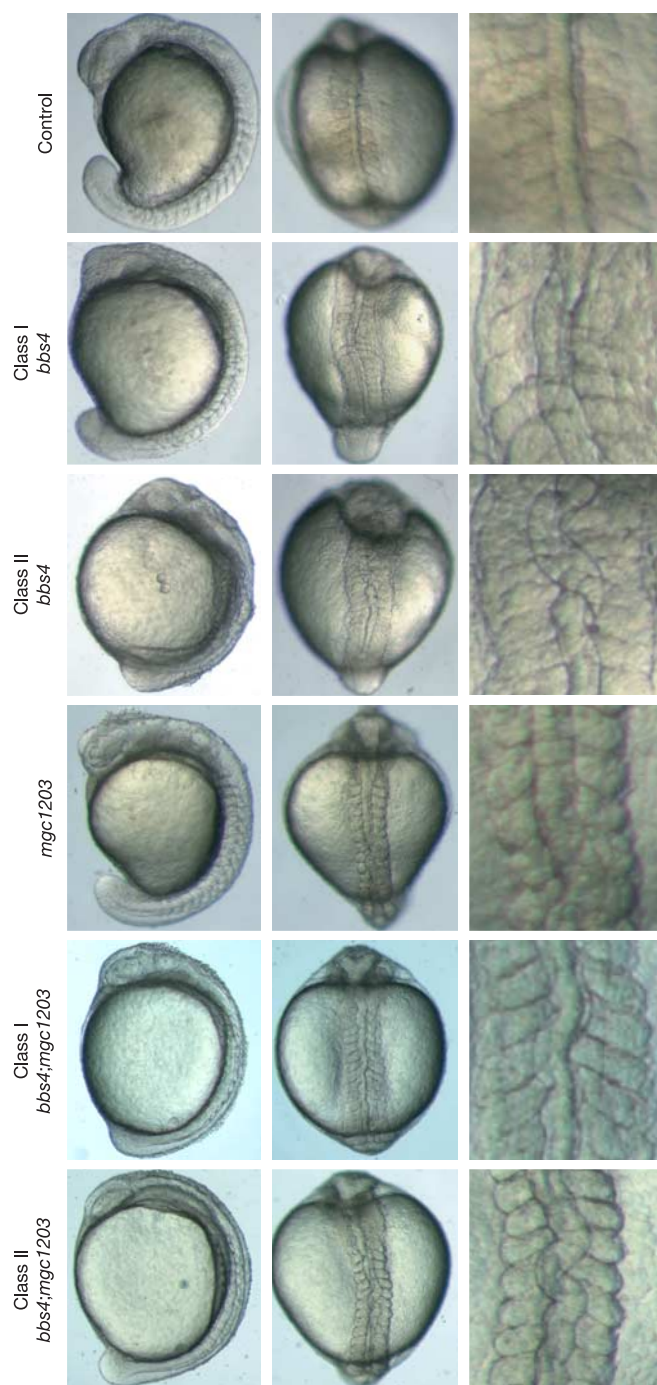


Figure 3 | Genetic interaction of *mgc1203* with *bbs4* in zebrafish. Live embryos showing characteristic class I (mild) or class II (severe) phenotypes of animals injected either with a *bbs4* morpholino alone, a high dose of *mgc1203* morpholino, or injected with low dose of both morpholinos. Side views (left panels) illustrate the shortening of the body axis and, for *bbs4* morphants, dorsal thinning. Note the detachment of cells along the rostrocaudal axis. Dorsal views (middle panels; magnified in right panels) show somitic and notochordal defects in the morphants. Original magnification: $\times 6.6$ (left and middle panels); $\times 26.4$ (right panels).

our example, the penetrance modification is contributed by an allele at a discrete, yet functionally related, locus.

Establishing the effect of epistatic variants can be challenging, given that hypomorphic mutations are typically in Hardy–Weinberg equilibrium in the population. Our studies highlight the usefulness of combinatorial approaches for the dissection of such phenomena

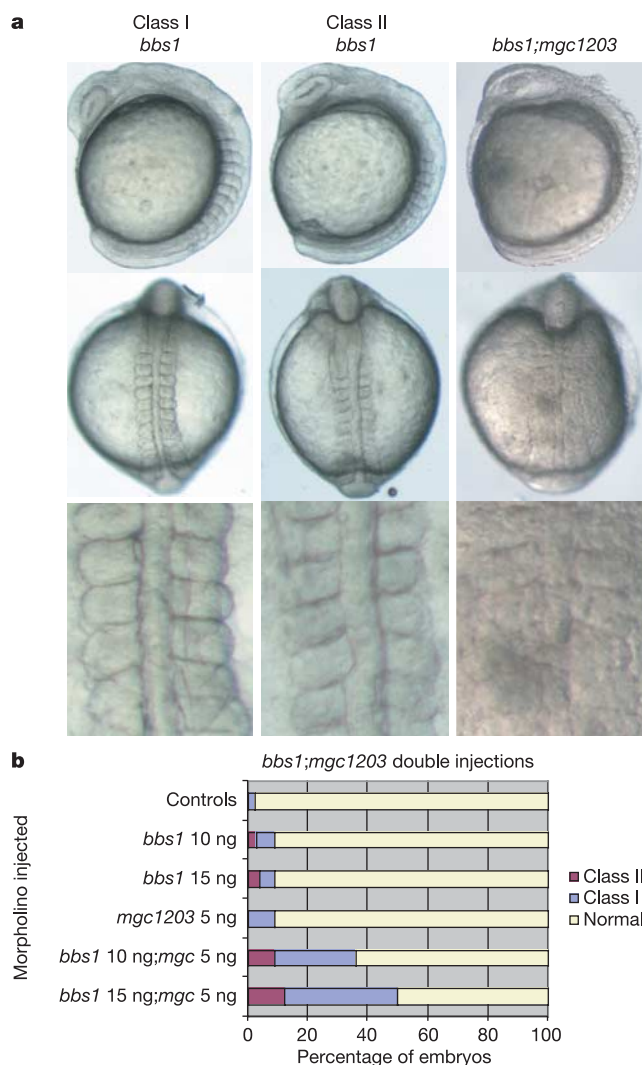


Figure 4 | Potent interaction between *mgc1203* and *bbs1*. **a**, Live zebrafish images showing a range of embryo phenotypes for the *bbs1* morpholino either alone, or in combination with the *mgc1203* morpholino. Note the particularly poor somitic definition of *bbs1*;*mgc1203* double morphants. **b**, Plot of the prevalence of phenotypes arising from the various injections. Original magnification: $\times 6.6$ (top and middle panels); $\times 26.4$ (bottom panels).

and exemplify how human genetics can lead to the functional investigation of phenotypic synthesis and further validation of previously established theoretical models^{29,30}. Improved understanding of the relationships between the components of a given molecular circuitry will be important for improving the ability of genotypic information to predict the phenotype in both oligogenic and complex traits.

METHODS

Yeast two-hybrid assay and immunoprecipitations. We performed RT–PCR on human RNA to amplify the open reading frame of *BBS-1*, -2, -4, -5, -6, -7 and -8, which were then cloned into the pSOS vector (Stratagene) and pCMV–Myc and pCMV–HA vectors (Clontech). To identify putative BBS4 interacting proteins, we used pSOS–BBS4 as bait in the Cytotrap yeast two-hybrid assay (Stratagene) with a fetal brain library as prey as described¹⁹. To confirm putative interactors, we assessed their ability to bind the BBS proteins by co-immunoprecipitation (co-IP) in an *in vitro* mammalian system as described¹⁰ (Supplementary Methods).

Mammalian cell culture and microscopy. HEK293T, HeLa and IMCD3 cells used for transient transfection, co-immunoprecipitations and immunohistochemistry were grown in standard conditions (Supplementary Methods). For fluorescence microscopy, cells grown on glass coverslips were treated as described¹⁹ (Supplementary Methods).

BBS patients and mutational analyses. Subject examination, clinical data, and samples were obtained under informed consent, and a diagnosis of BBS was secured according to established criteria. To sequence the complete open reading frame and exon–intron boundaries of *MGC1203*, we designed PCR primers as described¹² (Supplementary Methods). Primer sequences are available upon request. To analyse the resulting sequence reads, we used the Sequencher sequence alignment program (Gene Codes).

MGC1203 minigene construction and analysis. We amplified a 2.5-kb genomic fragment from exon 2 to exon 4 of *MGC1203* by PCR with TaKaRa LA Taq DNA polymerase and cloned it into the pcDNA3.1(+) vector (Invitrogen). All constructs were sequenced verified across their entire length by bidirectional dye-primer sequencing. Transfections were carried out in HEK293T cells with the Calcium Phosphate kit (Invitrogen) as described above, and cells were collected 24 h later. Total RNA was extracted and first-strand cDNA was prepared as described earlier. First-strand cDNA was then used as template in quantitative real-time PCR assays. Assays for the *MGC1203* transcript and the 5-bp deleted product, as well as 18S RNA and neomycin, were custom made by ABI using FAM-MGB probes (Applied Biosystems). Real-time PCR reactions were carried out using a 7900HT Sequence Detection System (Applied Biosystems).

Morpholinos and embryo manipulations. Translational and splice morpholinos against *bbs4*, *bbs6* and *mgc1203*, as well as a control morpholino, were designed by and obtained from Gene Tools (see the Supplementary Methods). One nanolitre of diluted morpholino was injected into wild-type zebrafish embryos at the 1-to-2-cell stage. Injected embryos were observed for 24–30 h and scored. For RNA rescue experiments, *bbs4*, *bbs6* and *mgc1203* mRNA was transcribed *in vitro* using the SP6 mMessage mMachine kit (Ambion). Morphant embryos were classified into two graded phenotypes depending on the relative severity compared to age-matched controls from the same clutch (see the Supplementary Methods for a description).

Received 4 August; accepted 25 October 2005.

Published online 4 December 2005.

1. Badano, J. L. & Katsanis, N. Beyond Mendel: an evolving view of human genetic disease transmission. *Nature Rev. Genet.* **3**, 779–789 (2002).
2. Badano, J. L. *et al.* Identification of a novel Bardet–Biedl syndrome protein, BBS7, that shares structural features with BBS1 and BBS2. *Am. J. Hum. Genet.* **72**, 650–658 (2003).
3. Badano, J. L. *et al.* Heterozygous mutations in *BBS1*, *BBS2* and *BBS6* have a potential epistatic effect on Bardet–Biedl patients with two mutations at a second BBS locus. *Hum. Mol. Genet.* **12**, 1651–1659 (2003).
4. Beales, P. L. *et al.* Genetic interaction of *BBS1* mutations with alleles at other *BBS* loci can result in non-Mendelian Bardet–Biedl syndrome. *Am. J. Hum. Genet.* **72**, 1187–1199 (2003).
5. Fan, Y. *et al.* Mutations in a member of the Ras superfamily of small GTP-binding proteins causes Bardet–Biedl syndrome. *Nature Genet.* **36**, 989–993 (2004).
6. Fauser, S., Munz, M. & Besch, D. Further support for digenic inheritance in Bardet–Biedl syndrome. *J. Med. Genet.* **40**, e104 (2003).
7. Katsanis, N. *et al.* Triallelic inheritance in Bardet–Biedl syndrome, a mendelian recessive disorder. *Science* **293**, 2256–2259 (2001).
8. Katsanis, N. *et al.* *BBS4* is a minor contributor to Bardet–Biedl syndrome and may also participate in triallelic inheritance. *Am. J. Hum. Genet.* **71**, 22–29 (2002).
9. Li, J. B. *et al.* Comparative genomic identification of conserved flagellar and basal body proteins that includes a novel gene for Bardet–Biedl syndrome. *Cell* **117**, 541–552 (2004).
10. Ansley, S. J. *et al.* Basal body dysfunction is a likely cause of pleiotropic Bardet–Biedl syndrome. *Nature* **425**, 628–633 (2003).
11. Chiang, A. P. *et al.* Comparative genomic analysis identifies an ADP-ribosylation factor-like gene as the cause of Bardet–Biedl syndrome (BBS3). *Am. J. Hum. Genet.* **75**, 475–484 (2004).
12. Katsanis, N. *et al.* Mutations in *MKKS* cause obesity, retinal dystrophy and renal malformations associated with Bardet–Biedl syndrome. *Nature Genet.* **26**, 67–70 (2000).
13. Mykytyn, K. *et al.* Identification of the gene that, when mutated, causes the human obesity syndrome BBS4. *Nature Genet.* **28**, 188–191 (2001).
14. Mykytyn, K. *et al.* Identification of the gene (*BBS1*) most commonly involved in Bardet–Biedl syndrome, a complex human obesity syndrome. *Nature Genet.* **31**, 435–438 (2002).
15. Nishimura, D. Y. *et al.* Positional cloning of a novel gene on chromosome 16q causing Bardet–Biedl syndrome (BBS2). *Hum. Mol. Genet.* **10**, 865–874 (2001).
16. Slavotinek, A. M. *et al.* Mutations in *MKKS* cause Bardet–Biedl syndrome. *Nature Genet.* **26**, 15–16 (2000).
17. Badano, J. L., Teslovich, T. M. & Katsanis, N. The centrosome in human genetic disease. *Nature Rev. Genet.* **6**, 194–205 (2005).
18. Katsanis, N. The oligogenic properties of Bardet–Biedl syndrome. *Hum. Mol. Genet.* **13**, R65–R71 (2004).
19. Kim, J. C. *et al.* The Bardet–Biedl protein BBS4 targets cargo to the pericentriolar region and is required for microtubule anchoring and cell cycle progression. *Nature Genet.* **36**, 462–470 (2004).
20. Kim, J. C. *et al.* *MKKS/BBS6*, a divergent chaperonin-like protein linked to the obesity disorder Bardet–Biedl syndrome, is a novel centrosomal component required for cytokinesis. *J. Cell Sci.* **118**, 1007–1020 (2005).
21. Cartegni, L., Wang, J., Zhu, Z., Zhang, N. Q. & Krainer, A. R. ESEfinder: a web resource to identify exonic splice enhancers. *Nucleic Acids Res.* **31**, 3568–3571 (2003).
22. Fu, X. & Maniatis, T. Isolation of a complementary DNA that encodes the mammalian splicing factor SC35. *Science* **256**, 533–538 (1992).
23. Gabut, M. *et al.* The SR protein SC35 is responsible for aberrant splicing of the *E1 α* pyruvate dehydrogenase mRNA in a case of mental retardation with lactic acidosis. *Mol. Cell. Biol.* **25**, 3286–3294 (2005).
24. Li, X. & Manley, J. L. Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability. *Cell* **122**, 365–378 (2005).
25. Lele, Z. *et al.* *parachute/n-cadherin* is required for morphogenesis and maintained integrity of the zebrafish neural tube. *Development* **129**, 3281–3294 (2002).
26. Ross, A. J. *et al.* Disruption of Bardet–Biedl syndrome ciliary proteins perturbs planar cell polarity in vertebrates. *Nature Genet.* **37**, 1135–1140 (2005).
27. Simons, M. *et al.* *Inversin*, the gene product mutated in nephronophthisis type II, functions as a molecular switch between Wnt signalling pathways. *Nature Genet.* **37**, 537–543 (2005).
28. Gouya, L. *et al.* The penetrance of dominant erythropoietic protoporphyria is modulated by expression of wildtype *FECH*. *Nature Genet.* **30**, 27–28 (2002).
29. Hartman, J., Gavrik, B. & Hartwell, L. Principles for the buffering of genetic variation. *Science* **291**, 1001–1004 (2001).
30. Kacser, H. & Burns, J. A. The molecular basis of dominance. *Genetics* **97**, 639–666 (1981).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the BBS patients and their families for their continued support and encouragement. We also thank J. Gerdes and A. McCallion for their thoughtful comments on the manuscript. This work was supported by grants from the National Institute of Child Health and Development (N.K.), the National Institute of Diabetes, Digestive and Kidney disorders (N.K.), the National Institute for Arthritis and Musculoskeletal disorders (S.F.), the Polycystic Kidney Disease Foundation (J.L.B. and N.K.), and the Medical Research Council (P.L.B.). R.A.L. is a Senior Scientific Investigator of Research to Prevent Blindness. P.L.B. is a Senior Wellcome Trust Fellow. H.C.D. is an Investigator of the Howard Hughes Medical Institute.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to N.K. (katsanis@jhmi.edu).

DNA sequence and analysis of human chromosome 8

Chad Nusbaum¹, Tarjei S. Mikkelsen¹, Michael C. Zody¹, Shuichi Asakawa², Stefan Taudien³, Manuel Garber¹, Chinnappa D. Kodira¹, Mary G. Schueler⁴, Atsushi Shimizu², Charles A. Whittaker^{1†}, Jean L. Chang¹, Christina A. Cuomo¹, Ken Dewar^{1†}, Michael G. FitzGerald¹, Xiaoping Yang¹, Nicole R. Allen¹, Scott Anderson¹, Teruyo Asakawa², Karin Blechschmidt³, Toby Bloom¹, Mark L. Borowsky¹, Jonathan Butler¹, April Cook¹, Benjamin Corum¹, Kurt DeArellano¹, David DeCaprio¹, Kathleen T. Dooley¹, Lester Dorris III¹, Reinhard Engels¹, Gernot Glöckner³, Nabil Hafez¹, Daniel S. Hagopian¹, Jennifer L. Hall¹, Sabine K. Ishikawa², David B. Jaffe¹, Asha Kamat¹, Jun Kudoh², Rüdiger Lehmann³, Tashi Lokitsang¹, Pendexter Macdonald¹, John E. Major¹, Charles D. Matthews¹, Evan Mauceli¹, Uwe Menzel^{3†}, Atanas H. Mihalev¹, Shinsei Minoshima^{2†}, Yuji Murayama², Jerome W. Naylor¹, Robert Nicol¹, Cindy Nguyen¹, Sinéad B. O'Leary¹, Keith O'Neill¹, Stephen C. J. Parker^{1†}, Andreas Polley^{3†}, Christina K. Raymond¹, Kathrin Reichwald^{3†}, Joseph Rodriguez¹, Takashi Sasaki², Markus Schilhabel³, Roman Siddiqui³, Cherylyn L. Smith¹, Tam P. Sneddon⁵, Jessica A. Talamas¹, Pema Tenzin¹, Kerri Topham¹, Vijay Venkataraman¹, Gaiping Wen^{3†}, Satoru Yamazaki², Sarah K. Young¹, Qiandong Zeng¹, Andrew R. Zimmer¹, Andre Rosenthal^{3†}, Bruce W. Birren¹, Matthias Platzer³, Nobuyoshi Shimizu² & Eric S. Lander¹

The International Human Genome Sequencing Consortium (IHGSC) recently completed a sequence of the human genome¹. As part of this project, we have focused on chromosome 8. Although some chromosomes exhibit extreme characteristics in terms of length, gene content, repeat content and fraction segmentally duplicated, chromosome 8 is distinctly typical in character, being very close to the genome median in each of these aspects. This work describes a finished sequence and gene catalogue for the chromosome, which represents just over 5% of the euchromatic human genome. A unique feature of the chromosome is a vast region of ~15 megabases on distal 8p that appears to have a strikingly high mutation rate, which has accelerated in the hominids relative to other sequenced mammals. This fast-evolving region contains a number of genes related to innate immunity and the nervous system, including loci that appear to be under positive selection²—these include the major defensin (DEF) gene cluster^{3,4} and *MCPHI*^{5,6}, a gene that may have contributed to the evolution of expanded brain size in the great apes. The data from chromosome 8 should allow a better understanding of both normal and disease biology and genome evolution.

The finished sequence of chromosome 8 contains 145,556,489 bases and is interrupted by only four euchromatic gaps, one gap at the 8p telomere and one gap containing the centromeric heterochromatin (Fig. 1 and Supplementary Table S1). These gaps are refractory to current cloning and mapping technology. The estimated total size of the euchromatic gaps is 427 kilobases (kb), based

on direct sizing of three gaps and estimation of the remaining two gaps at the genome-wide average of ~100 kb each. This corresponds to ~0.3% of the euchromatic length of the chromosome, similar to the genome average^{1,7–11}. In all, 182.3 megabases (Mb) of finished sequence were generated by the Broad Institute of MIT and Harvard (formerly Whitehead Institute/MIT Center for Genome Research (WICGR)), 27.9 Mb by Keio University School of Medicine, 8.4 Mb by the Institute of Molecular Biotechnology in Jena, and 5.8 Mb by 10 other groups (Supplementary Tables S2 and S3). These sequences (which include overlap) were combined to yield the finished path (see Methods).

We assessed the local accuracy of the clone path by aligning paired-end sequences from a human Fosmid library (WIBR2, representing ×10 physical coverage) to the finished sequence⁷. Errors in the clone path were detected by identifying discrepancies between the predicted and observed distances between Fosmid ends⁷. This revealed two deleted clones, which were replaced. Finally, an independent quality assessment exercise commissioned by NHGRI estimated the accuracy of the finished sequence at less than 1 error in 100,000 bases¹² (J. Schmutz, personal communication).

Several analyses support the idea that nearly the entire euchromatic region of chromosome 8 is present and accurately represented. From the well-curated RefSeq¹³ data set 681 transcripts (from 573 unique genes) mapped to chromosome 8. All but one of these are present and complete in the finished sequence. The finished sequence shows excellent co-linearity with the genetic map¹⁴ (Supplementary

¹Broad Institute of MIT and Harvard, 320 Charles St, Cambridge, Massachusetts 02141, USA. ²Department of Molecular Biology, Keio University School of Medicine, 35 Shinanomachi, Shinjuku-ku, Tokyo 160-8582, Japan. ³Genome Analysis, Institute of Molecular Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany. ⁴National Human Genome Research Institute, National Institutes of Health, 50 South Drive Rm 5529, Bethesda, Maryland 20982, USA. ⁵HUGO Gene Nomenclature Committee (HGNC), The Galton Laboratory, Department of Biology, University College London, Wolfson House, 4 Stephenson Way, London NW1 2HE, UK. [†]Present addresses: MIT Center for Cancer Research, 77 Massachusetts Avenue E18-570, Cambridge, Massachusetts 02139, USA (C.A.W.); McGill University and Genome Quebec Innovation Centre, Montreal, Quebec H3A 1A4, Canada (K.D.); Department of Genetics and Pathology, Uppsala University, SE-751 85 Uppsala, Sweden (U.M.); Photon Medical Research Center, Hamamatsu University School of Medicine, Handayama, Hamamatsu, Shizuoka 431-3192, Japan (S.M.); Boston University Bioinformatics and Systems Biology Program, 24 Cummington St, Boston, Massachusetts 02215, USA (S.C.J.P.); TraitGenetics GmbH, Am Schwabeplan 1b, 06466 Gatersleben, Germany (A.P.); University Clinic for Child and Adolescent Psychiatry, University of Duisburg-Essen, Virchowstr. 174, 45147 Essen, Germany (K.R.); GSF-Forschungszentrum für Umwelt und Gesundheit, Ingolstädter Landstraße 1, 85674 Neuherberg, Germany (G.W.); Signature Diagnostics AG, Voltairerweg 4B, 14469 Potsdam, Germany (A.R.).

Fig. S1). Among 247 sequence-based genetic markers (Supplementary Table S4) there are six discrepancies. One discrepancy consists of eight markers and spans a region in 8p23 known to be the site of a polymorphic inversion in the human population^{15,16} (see below). Five discrepancies each consist of single markers out of order by one position; all occur in small regions where the genetic map shows no recombination in one of the two sexes (Supplementary Table S4). The sequence also shows good agreement with the radiation hybrid (RH) map¹⁷ (Supplementary Table S5).

We produced a manually curated gene catalogue, containing 793 gene loci and 301 pseudogene loci (see Methods). The catalogue includes all previously known genes on chromosome 8 (Table 1). According to the Hawk2 categorization scheme¹⁸, there are 614 'known' genes, 109 'novel CDS', 43 'novel transcripts', 14 'putatives' and 13 'gene fragments'. The small set of novel and putative categories were annotated by spliced expressed sequence tag (EST) evidence only; some 'putative novel' loci may prove to be pseudogenes. Comparison of manual annotation performed at the Broad Institute of MIT and Harvard to manual annotation for specific regions done at Jena and Keio indicated that they were largely the same, and that virtually all differences could be attributable to edge effects (see Supplementary Information).

Full-length transcripts of known genes contain an average of 9.9 exons, comparable to recently published reports^{8–11,19}, have an average length of 3,056 base pairs (bp), and internal exons have an average length of 155 bp. There is evidence of extensive alternate splicing. Gene loci have an average of 4.1 distinct transcripts, with 63% having at least two transcripts, values that are similar to recent reports^{8,9,11,20}. Of the 301 pseudogenes on chromosome 8, ~84% are processed pseudogenes arising from retrotransposition; the remaining 16% are unprocessed. We also identified 13 tRNA genes (Supplementary Table S6). Examples of genes that represent extremes from these averages are described in Supplementary Information.

Several aspects of the genome landscape are notable. The overall gene density is 5.6 genes Mb⁻¹, below the genome average of ~10 genes Mb⁻¹. Gene distribution is highly heterogeneous, with 44 gene deserts (500 kb without a coding gene, Supplementary Table S7) that together comprise 41.9 Mb or ~29% the total length. The overall G+C content is 39.2%, but varies substantially across the chromosome (Fig. 1). Nearly half of the chromosome is composed of repeat sequences, with transposable element fossils comprising 44.5%, low complexity sequence (including simple sequence repeats and satellite sequences) comprising 1.8%, and segmental duplications comprising ~2.1% (with interchromosomal

and intrachromosomal duplications at ~1.5% each, with some sequence included in both categories) (E. Eichler and X. She, personal communication).

Chromosome 8 is the first human autosome and one of only two chromosomes (the other being chromosome X²⁰) for which sequences span the entire pericentromeric region. The regions on both arms stretch from unique euchromatin through pericentromeric satellites and into the higher-order alpha-satellite array (Fig. 2). Three variant higher-order repeat units populate the chromosome 8 higher-order array, *D8Z2* (ref. 21 and Supplementary Information). The proximal termini of both the 8p and 8q sequence contigs are comprised of nine copies of the 1.9-kb unit. The p and q arm higher-order units are highly identical to each other (96–98%) and occur in the same head-to-tail orientation, indicating that these sequences sample the edges of the chromosome 8-specific array. Analysis of the finished pericentromeric sequence of chromosome 8 is essential to test and further develop primate centromere evolution hypotheses using an autosomal model.

The most striking feature on chromosome 8 emerges from evolutionary and population genetic comparisons (Fig. 3). The most distal 15 Mb on chromosome 8p show an extremely high divergence between human and chimpanzee (0.021 substitutions per site, 4.0 s.d. above the mean of 0.012). The region also shows a strikingly high polymorphism rate in the human population (0.0018, 3.2 s.d. above the mean of 0.0010). The peak divergence reaches 0.032 (8.6 s.d.), and diversity 0.0028 (7.1 s.d.), across a 1-Mb region (3.3–4.3 Mb) overlapping the *CSMD1* gene. This is the highest divergence level seen across all autosomes and chromosome X. Only regions of chromosome Y may be more rapidly diverging, driven by the high mutation rate in the male germ line. We excluded trivial explanations for this observation, such as unresolved segmental duplications (Supplementary Information). Diversity is also locally high in the chimpanzee, although the data are more limited.

The high rate of divergence and diversity at distal 8p might reflect either an extraordinary mutation rate or population genetic history. The latter alternative would require an unusually long coalescence time to the most recent common ancestor over a very large region; this would be remarkable inasmuch as local coalescence times tend to be correlated over short distances, as the correlation falls below 0.5 within 20 kb (ref. 22). We sought to resolve the issue by examining the divergence rates with more distant mammalian species, where the impact of population genetic history should be negligible.

Comparison of ancestral interspersed repeats in the human, dog²³

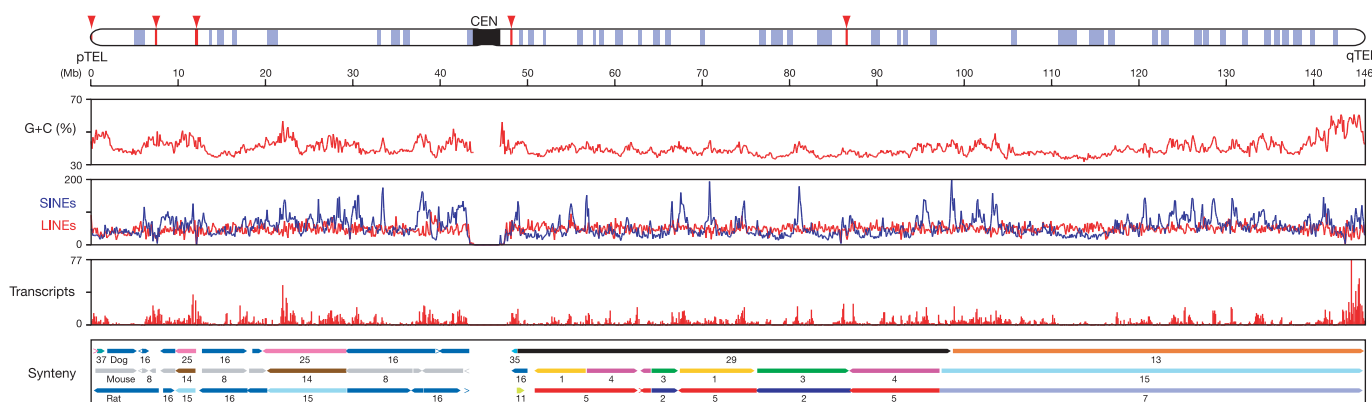


Figure 1 | Overview of human chromosome 8. The features are addressed in the order of top to bottom. In the cartoon, blue shading indicates gene deserts (≥ 500 kb with no transcript, Supplementary Table S7); telomeres (pTEL and qTEL), the centromere (CEN) and euchromatic sequence gaps (red lines) are indicated. The following features are represented in discrete windows of 100 kb: G+C content (on a scale from 30–70%); densities of

LINEs (long interspersed nucleotide elements; red) and SINEs (short interspersed nucleotide elements; blue); and densities of transcripts (all are counts of elements). The box at the bottom shows blocks of conserved syntenicity (100-kb resolution) with dog, mouse and rat as determined for this work. Chromosomes are numbered, and are coloured arbitrarily for ease of distinction.

Table 1 | Chromosome 8 gene content

Category	Gene number	Gene percentage	Gene length (bp)*	Number of alternative transcripts	Transcript length (bp)†	Number of exons per transcript‡	Internal exon length (bp)§	Intron length (bp)	CpG-5' association¶
Known gene	614	77	81,744	4.1	3,056	9.9	155 (n = 5,725)	9,630 (n = 7,710)	77
Novel transcript	43	5	96,268	1.8	1,116	3.8	146 (n = 127)	27,207 (n = 248)	42
Putative gene	14	2	45,433	1.2	714	2.4	123 (n = 21)	23,787 (n = 57)	36
Novel CDS	109	14	21,890	1.9	1,142	5.0	138 (n = 487)	5,103 (n = 625)	29
Gene fragment	13	2	648	1.0	648	1.0	-	-	8
Total	793	-	72,334	-	-	-	-	-	-
Pseudogene	301	28	1,334	1.0	875	1.3	195 (n = 50)	1,430 (n = 97)	5

* Average chromosomal distance from beginning of 5'-most exon to 3'-most exon in all transcripts in a gene.
† Average length summed across the footprint of all exons in all transcripts in a gene—total exon space per gene.
‡ Average number of exons in transcripts. Exons common to different transcripts were counted once per transcript.
§ Average length of exons using the footprint of all non-terminal exons of all transcripts in a gene. Unique overlapping exons or contained exons are counted separately, making this an average length of unique exons in a gene.
|| Average length of unique introns in a gene. In the case of exon skipping, both the shorter and longer versions of the overlapping introns were counted towards the average.
¶ Percentage of genes with a transcript having a CpG island (as assessed by FirstEF) within -2 kb and +1 kb of transcription start.

and mouse²⁴ genomes reveals that the region exhibits above-average lineage-specific divergence rates on all three lineages across 100 million years of evolution, but that the rate is the most elevated relative to the genome-wide mean in the lineage leading to humans. The greatest elevation is seen in the most distal 6 Mb of 8p, where the ancestral interspersed repeat divergence rates in the orthologous sequences have been 0.19 (3.3 s.d. above the mean of 0.14) on the human lineage and 0.41 (1.0 s.d. above the mean of 0.38) in the mouse lineage since the primate-rodent split, and 0.24 (1.9 s.d. above the mean of 0.20) in the dog lineage since the divergence from the common boreo-eutherian ancestor.

The biological basis for the apparently high mutation rate is unclear. Three major factors have been associated with high mutation rates in the human genome: proximity to telomeres, high

recombination rate and high A+T content^{25,26}. The region on chromosome 8p has all three factors. The mean sex-averaged recombination rate across the first 6 Mb is 2.7 cM Mb⁻¹, with a 1-Mb window peak of 3.5, as compared to the genome-wide average of 1.2. The region from 2.5–6 Mb is 62% A+T, as compared to a genome-wide average of 59%. It is unusual in this regard, because subtelomeric regions with high recombination rates are typically (A+T)-poor. Notably, the region is not subtelomeric in the mouse, where the lowest rate elevation is observed.

The distal region on chromosome 8p also contains at least two loci that appear to be undergoing positive selection (Fig. 3). The first locus is the major cluster of defensin genes, which lies within the region of high mutation (5.5–7.5 Mb), although ~2.5 Mb from the peak. The defensin genes express small cationic antimicrobial

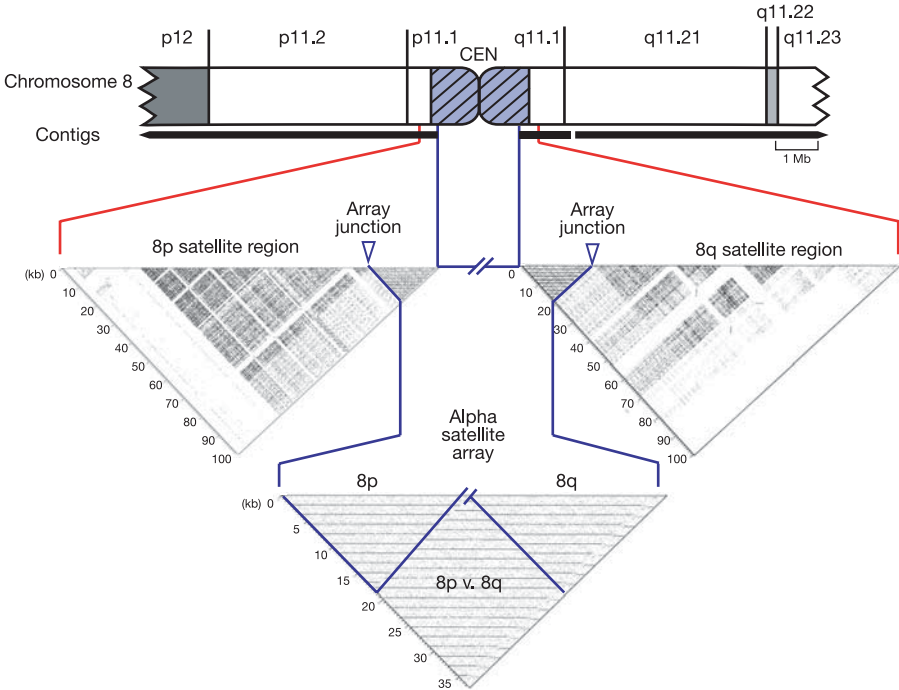


Figure 2 | 8p and 8q pericentromeric contigs extend into chromosome 8-specific higher-order alpha satellite, D8Z2. The pericentromeric region of chromosome 8 is shown as a truncated ideogram with the extent of sequence coverage shown below by black bars. Dotter plots show self-alignments of the most proximal ~100 kb from each arm including ~36 kb of the chromosome-specific alpha satellite array (D8Z2). Junctions between the arm-specific satellite region and D8Z2 are marked with blue arrows. Dark blocks indicate the highly repetitive nature of the satellite region and mark similarity between monomers within each satellite family. Gaps in the

dark blocks occur where interspersed elements (LINEs, SINEs and long terminal repeats) interrupt the satellite sequences. In the alpha satellite array dotter plot (bottom), D8Z2 from 8p (~18 kb) is joined with that of 8q (~18 kb). The plot reveals the periodic nature of the centromeric, higher-order alpha satellite array with black horizontal lines indicating near identity of sequences spaced at ~1.9-kb intervals. The regions outlined in blue are self-self alignments ('8p' and '8q'), whereas the remaining rectangular region of the plot is an alignment of 8p versus 8q D8Z2.

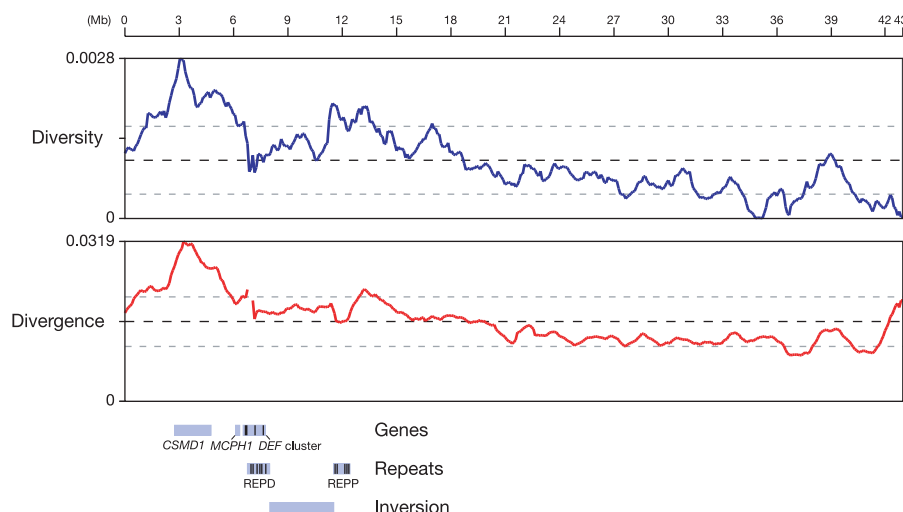


Figure 3 | Diversity and divergence on 8p. Coloured lines indicate the distribution of human diversity (blue) and human–chimpanzee divergence (red). Values of genome averages and of 2 standard deviations from the means are indicated (dark and light dashed lines, respectively). Features mentioned in the text are indicated in the bottom panel, including genes, two low copy repeats (LCRs) and the common 8p23 inversion. Vertical ticks

in the LCR boxes indicate olfactory receptor genes or pseudogenes, and vertical ticks in the DEF cluster boxes represent individual defensin (DEF) genes. There is a discontinuity in the divergence plot from 6.98 to 8.13 Mb. This region, corresponding to the REPD repeat, is also highly duplicated in the chimpanzee, making it impossible to align sequence with high enough confidence to call divergence.

peptides crucial to the innate immune response²⁷. Studies^{2,3} have suggested that defensins have been under positive selection, with a high ratio of non-synonymous to synonymous changes detected in the mature peptide coding exon. Moreover, gene and segmental duplication within the cluster have led to extensive copy number^{28,29} and haplotype³⁰ polymorphism within and across populations, which are thought to influence variation in disease susceptibility and contribute to ongoing adaptive evolution in both the human and chimpanzee species. The second locus showing positive selection is *MCPH1*, mutations in which cause microcephaly (Online Mendelian Inheritance in Man (OMIM): 251200); there is clear evidence of accelerated non-synonymous divergence correlating with the expansion of brain size throughout the lineage from simian ancestors to the human and chimpanzee^{4,5}.

To investigate the diversity of copy number in the defensin clusters, we resequenced several dozen polymerase chain reaction (PCR) products from representative intervals from *DEFB105A* (beta-defensin cluster) and *DEFA1* (alpha-defensin cluster) in 14 chimpanzees, 1 gibbon, 1 macaque and 4 breeds of dog (see Methods and Supplementary Information). In all species studied, the gene family has multiple members, and the members are more similar within a species than across species. Thus, the defensin clusters have either independently duplicated in each species or have undergone gene conversion events within species.

Finally, we note that the majority of the genes in the region of high divergence in distal 8p play important roles in development or signalling in the nervous system. Notably, the extremely large *CSMD1* gene, which lies at the peak of divergence and diversity, is widely expressed in brain tissues. High regional mutation rates and positive selection are generally assumed to be distinct, but it is possible that the former may facilitate the latter by increasing the rate of appearance of potentially advantageous single, or interacting, alleles (see also ref. 31). It is intriguing to speculate whether the accelerated divergence rate of this region has contributed to the rapid expansion and evolution of the primate brain.

METHODS

See Supplementary Information for details on clone path building, generation of sequence map, sizing of gaps and gene annotation. The final version of the clone path is available in AGP format (see <http://www.ncbi.nlm.nih.gov/genome/guide/glossary.htm>) at <http://www.broad.mit.edu/tools/data/data-human.html>.

Gene amplification and sequencing. TBLASTN (<http://www.ncbi.nlm.nih.gov/BLAST>) was used to identify *DEFB105* and *DEFA1* orthologues in 16 chimpanzees, 1 gibbon, 1 macaque and 4 dog breeds (akita, golden retriever, greyhound and mastiff). PCR primers for gene amplification were designed using Primer3 (<http://frodo.wi.mit.edu/primer3>) based on the species reference sequence. Human and macaque primers were used for gibbon. Amplified products were cloned, and for each individual/gene combination, 48 or 96 clones were sequenced.

Haplotype analysis. Neighbourhood Quality Standard³² (NQS) scores were computed for all sequenced products using the published constraints³². Reads were trimmed to the first and last three consecutive NQS bases, and aligned to the reference sequence using PatternHunter (<http://www.bioinformaticsolutions.com>). Multiple sequence alignments were built from the pairwise alignments and inspected to find SNPs that were: at NQS bases, supported by at least two reads, and in a ten base window where not more than two other variations were observed. To minimize false positives due to errors during PCR amplification, we restricted our analysis to haplotypes that differed in >3 bases.

Received 5 August; accepted 6 October 2005.

1. International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
2. Vallender, E. J. & Lahn, B. T. Positive selection on the human genome. *Hum. Mol. Genet.* **13** (suppl. 2), R245–R254 (2004).
3. Maxwell, A. I., Morrison, G. M. & Dorin, J. R. Rapid sequence divergence in mammalian β -defensins by adaptive evolution. *Mol. Immunol.* **40**, 413–421 (2003).
4. Xiao, Y. *et al.* A genome-wide screen identifies a single β -defensin gene cluster in the chicken: implications for the origin and evolution of mammalian defensins. *BMC Genom.* **5**, 56 (2004).
5. Evans, P. D., Anderson, J. R., Vallender, E. J., Choi, S. S. & Lahn, B. T. Reconstructing the evolutionary history of microcephalin, a gene controlling human brain size. *Hum. Mol. Genet.* **13**, 1139–1145 (2004).
6. Evans, P. D. *et al.* Microcephalin, a gene regulating brain size, continues to evolve adaptively in humans. *Science* **309**, 1717–1720 (2005).
7. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature* **431**, 931–945 (2004).
8. Grimwood, J. *et al.* The DNA sequence and biology of human chromosome 19. *Nature* **428**, 529–535 (2004).
9. Deloukas, P. *et al.* The DNA sequence and comparative analysis of human chromosome 10. *Nature* **429**, 375–381 (2004).
10. Martin, J. *et al.* The sequence and analysis of duplication-rich human chromosome 16. *Nature* **432**, 988–994 (2004).
11. Nusbaum, C. *et al.* DNA sequence and analysis of human chromosome 18. *Nature* **437**, 551–555 (2005).
12. Schmutz, J. *et al.* Quality assessment of the human genome sequence. *Nature* **429**, 365–368 (2004).
13. Pruitt, K. D., Tatusova, T. & Maglott, D. R. NCBI Reference Sequence (RefSeq):

- a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* **33**, D501–D504 (2005).
14. Kong, A. *et al.* A high-resolution recombination map of the human genome. *Nature Genet.* **31**, 225–226 (2002).
 15. Giglio, S. *et al.* Olfactory receptor-gene clusters, genomic-inversion polymorphisms, and common chromosome rearrangements. *Am. J. Hum. Genet.* **68**, 874–883 (2001).
 16. Shimokawa, O. *et al.* Molecular characterization of inv dup del(8p): analysis of five cases. *Am. J. Med. Genet. A* **128**, 133–137 (2004).
 17. Deloukas, P. *et al.* A physical map of 30,000 genes. *Science* **282**, 744–746 (1998).
 18. Ashurst, J. L. *et al.* The Vertebrate Genome Annotation (Vega) database. *Nucleic Acids Res.* **33**, D459–D465 (2005).
 19. Hillier, L. W. *et al.* Generation and annotation of the DNA sequences of human chromosomes 2 and 4. *Nature* **434**, 724–731 (2005).
 20. Ross, M. T. *et al.* The DNA sequence of the human X chromosome. *Nature* **434**, 325–337 (2005).
 21. Ge, Y., Wagner, M. J., Siciliano, M. & Wells, D. E. Sequence, higher order repeat structure, and long-range organization of alpha satellite DNA specific to human chromosome 8. *Genomics* **13**, 585–593 (1992).
 22. Reich, D. E. *et al.* Human genome sequence variation and the influence of gene history, mutation and recombination. *Nature Genet.* **32**, 135–142 (2002).
 23. Lindblad-Toh, K. *et al.* Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* **438**, 803–819 (2005).
 24. Mouse Genome Sequencing Consortium, Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520–562 (2002).
 25. The Chimpanzee Sequencing and Analysis Consortium. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* **437**, 69–87 (2005).
 26. Hellmann, I. *et al.* Why do human diversity levels vary at a megabase scale? *Genome Res.* **15**, 1222–1231 (2005).
 27. Lehrer, R. I. Primate defensins. *Nature Rev. Microbiol.* **2**, 727–738 (2004).
 28. Hollox, E. J., Armour, J. A. & Barber, J. C. Extensive normal copy number variation of a β -defensin antimicrobial-gene cluster. *Am. J. Hum. Genet.* **73**, 591–600 (2003).
 29. Mars, W. M. *et al.* Inheritance of unequal numbers of the genes encoding the human neutrophil defensins HP-1 and HP-3. *J. Biol. Chem.* **270**, 30371–30376 (1995).
 30. Taudien, S. *et al.* Polymorphic segmental duplications at 8p23.1 challenge the determination of individual defensin gene repertoires and the assembly of a contiguous human reference sequence. *BMC Genom.* **5**, 92 (2004).
 31. Wyckoff, G. J., Malcom, C. M., Vallender, E. J. & Lahn, B. T. A highly unexpected strong correlation between fixation probability of nonsynonymous mutations and mutation rate. *Trends Genet.* **21**, 381–385 (2005).
 32. Altshuler, D. *et al.* An SNP map of the human genome generated by reduced representation shotgun sequencing. *Nature* **407**, 513–516 (2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank L. Gaffney for help with figures and tables; L. French and her group at the Sanger Institute for attempting fibre FISH analysis to size some clone gaps in the tiling path of chromosome 8; E. Eichler and X. She for sharing their data on segmental duplications; T. Furey for help with lists of genetic markers and placement of RefSeq genes; M. Kamal for assistance and advice with synteny analysis; and K. Lindblad-Toh for sharing data from the dog genome project. We also acknowledge the HUGO Gene Nomenclature Committee (S. Povey, chair) for assigning official gene symbols. We are deeply grateful to all the members, present and past, of the Genome Sequencing Platform of the Broad Institute (and Whitehead Center for Genome Research), Keio University School of Medicine and the Institute of Molecular Biology at Jena for their dedication and for the consistent high quality of their data that made this work possible. This work was supported by grants from the National Human Genome Research Institute, RIKEN, the 'Research for the Future' Program from the Japan Society for the Promotion of Science (JSPS), the Ministry of Education, Culture, Sports, Science and Technology of Japan (MEXT), the Federal German Ministry of Education, Research and Technology, and the Thüringer Kultusministerium.

Author Information Accession numbers for all clones contributing to the finished sequence of human chromosome 8 can be found in Supplementary Table S2. The updated human chromosome 8 sequence can be accessed through GenBank accession number NC_000008. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to C.N. (chad@broad.mit.edu) or N.S. (shimizu@dmf.med.keio.ac.jp).

LETTERS

Top-down gain control of the auditory space map by gaze control circuitry in the barn owl

Daniel E. Winkowski¹ & Eric I. Knudsen¹

High-level circuits in the brain that control the direction of gaze are intimately linked with the control of visual spatial attention^{1–5}. Immediately before an animal directs its gaze towards a stimulus, both psychophysical sensitivity to that visual stimulus and the responsiveness of high-order neurons in the cerebral cortex that represent the stimulus increase dramatically^{3,6,7}. Equivalent effects on behavioural sensitivity and neuronal responsiveness to visual stimuli result from focal electrical microstimulation of gaze control centres in monkeys^{8–11}. Whether the gaze control system modulates neuronal responsiveness in sensory modalities other than vision is unknown. Here we show that electrical microstimulation applied to gaze control circuitry in the forebrain of barn owls regulates the gain of midbrain auditory responses in an attention-like manner. When the forebrain circuit was activated, midbrain responses to auditory stimuli at the location encoded by the forebrain site were enhanced and spatial selectivity was sharpened. The same stimulation suppressed responses to auditory stimuli represented at other locations in the midbrain map. Such space-specific, top-down regulation of auditory responses by gaze control circuitry in the barn owl suggests that the central nervous system uses a common strategy for dynamically regulating sensory gain that applies across modalities, brain areas and classes of vertebrate species. This approach provides a path for discovering mechanisms that underlie top-down gain control in the central nervous system.

The possibility that gaze control circuitry mediates spatial attention for audition, as well as for vision, is suggested by psychophysical and neurophysiological data. For example, when human subjects are asked to report the relative location of a short-duration, pulsed white noise stimulus, reaction time decreases and accuracy improves when the auditory stimulus is located near the endpoint of a visually instructed eye saccade compared to when saccades are made in the opposite direction¹². In addition, responses of auditory neurons in the substantia nigra and caudate nucleus of monkeys increase when an auditory stimulus is the target of an upcoming eye saccade^{13,14}.

In this study, we tested the hypothesis that forebrain gaze control circuitry regulates auditory neural processing in a space-specific manner. The arcopallial gaze fields (AGF), a premotor region in the owl's forebrain, has a central role in the control of gaze direction¹⁵. This region is analogous to the frontal eye fields (FEF) in the mammalian frontal cortex. The AGF, like the FEF, projects in parallel to the deep layers of the optic tectum and to saccade-generating premotor neurons in the brainstem^{16,17}; electrical microstimulation in the AGF, like in the FEF, produces orienting movements of the eyes and head^{16,18}; and the AGF, like the FEF, mediates memory-guided saccades^{19,20}. Recently, the FEF in monkeys has also been shown to direct visual spatial attention and to regulate the responsiveness of extrastriate visual neurons in a space-specific manner^{9,10}. A similar role for the AGF has not previously been tested.

In our study, electrical microstimulation was applied to the AGF

while auditory responses were measured in the deep layers of the optic tectum, which contains a map of auditory space²¹. Auditory stimuli were presented via earphones so that spatial cues could be presented in randomly interleaved patterns and at high resolution, allowing us to characterize the spatial specificity of any gain changes. To simulate changes in stimulus location, we varied interaural timing differences (ITD), which indicate the horizontal position of a sound stimulus, and interaural level differences (ILD), which for barn owls indicate the vertical position of a stimulus²².

The auditory portion of the AGF contains neurons that are tuned for both ITD and ILD and, therefore, for space²³. Unlike in the optic tectum, where auditory space is mapped, in the AGF space is organized in a clustered representation in which neighbouring neurons encode a similar location, but neighbouring groups of neurons encode different, unpredictable locations²³. The stimulating microelectrode was positioned in the centre of an AGF cluster as follows. In dorsoventral electrode penetrations through the AGF, ITD and ILD tuning was measured at approximately 100- μ m intervals. When a cluster was found in which ITD and ILD tuning remained constant (± 15 μ s and ± 4 dB, respectively) for at least 300 μ m, the electrode was centred in the cluster and the binaural tuning at the stimulation site was measured. The threshold for eliciting a motor response was determined by passing balanced biphasic pulses (25-ms train, 200 Hz, 200- μ s phase duration) through the electrode and finding the lowest current level that elicited an eye saccade (mean 150 μ A, range 55–400 μ A). During the experiments, current amplitude was set to levels of less than half the threshold for eliciting a motor response and was never increased above 40 μ A.

To monitor auditory responses, a second microelectrode was introduced into the deep layers (layers 11–13) of the optic tectum, and binaural tuning at the recording site was measured. Best ITD and best ILD were defined as the weighted average of responses greater than half of the maximum response (henceforth referred to as 'half-max'). Tuning width was defined as the continuous range of ITD values that elicited a response greater than half-max. On the basis of that tuning, a stimulus set was programmed that sampled the entire ITD range for the site, at 10- or 20- μ s intervals (10–20 repetitions of each ITD value under each condition), while holding ILD constant at the site's best value. On half of the trials, the sound was immediately preceded by AGF microstimulation. All ITD values and conditions (with and without electrical microstimulation) were randomly interleaved. In sequential stimulus sets, the AGF microstimulation current was increased in small increments until an effect on auditory responses was observed or until the current level reached 40 μ A (maximum current tested).

Results from an experiment in which the binaural (spatial) tuning for ITD and ILD at the AGF and optic tectum sites were similar are shown in Fig. 1. For the AGF site, the best ITD was -10 μ s (leading in the left ear) and the best ILD was 1 dB (greater in the right ear). For the optic tectum site, the best ITD was -16 μ s (Δ ITD_{AGF-OT} = 6 μ s;

¹Department of Neurobiology, Stanford University, Stanford, California 94305, USA.

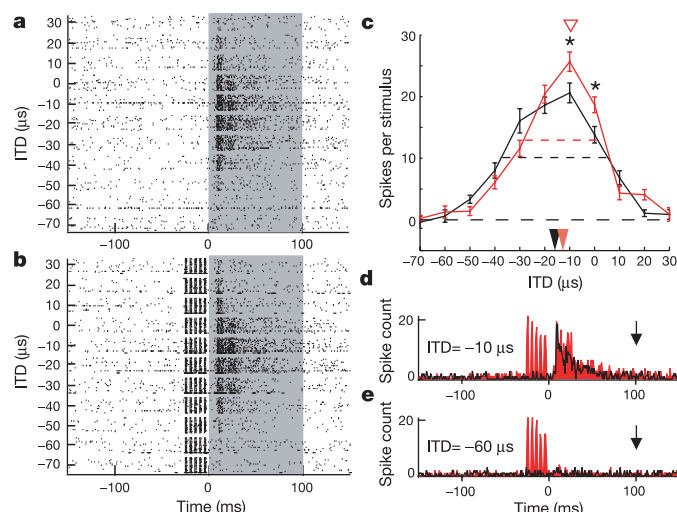


Figure 1 | Effect of electrical microstimulation in the AGF on auditory responses at an aligned site in the optic tectum. **a**, Raster display of optic tectum auditory responses without AGF stimulation. Ordinate shows stimulus ITD; abscissa shows time relative to stimulus onset; shaded area shows duration of auditory stimulus and the post-stimulus time-window used for data analysis. Best ITD $-16 \mu\text{s}$; tuning width at half-max $42 \mu\text{s}$. **b**, Same as **a**, except that electrical microstimulation was applied to the AGF site immediately before the onset of the auditory stimulus. Vertical lines show an electrical stimulation artefact. Best ITD of the optic tectum site with AGF stimulation $-12 \mu\text{s}$; tuning width at half-max $31 \mu\text{s}$. **c**, Mean neuronal response plotted as a function of stimulus ITD for trials with (red) and without (black) AGF stimulation. Error bars represent s.e.m. Asterisk, $P < 0.02$ (paired t -test). Solid arrowheads indicate best ITD with (red) and without (black) AGF microstimulation; open arrowhead indicates best ITD at the AGF site. **d**, Peristimulus time histograms of optic tectum responses to an auditory stimulus with an ITD of $-10 \mu\text{s}$ during trials with (red) and without (black) AGF stimulation. Arrow shows offset of auditory stimulus. **e**, Same as **d**, except histograms show responses to sound with an ITD of $-60 \mu\text{s}$ (outside the receptive field). Arrow shows offset of auditory stimulus.

$\sim 2^\circ$ azimuth) and the best ILD was 1 dB ($\Delta\text{ILD}_{\text{AGF-OT}} = 0 \text{ dB}$). Electrical microstimulation ($7 \mu\text{A}$, 25-ms train duration, 200 Hz; current to evoke eye saccades was $175 \mu\text{A}$) applied to the AGF site just before the onset of the sound stimulus increased the strength of the optic tectum auditory responses to ITDs within $10 \mu\text{s}$ of the best ITD (Fig. 1b–d) but had no effect on responses to sounds with ITDs outside of the receptive field (Fig. 1b, c, e). AGF microstimulation caused the best ITD of the optic tectum site to shift dynamically by $4 \mu\text{s}$ (best ITD with AGF microstimulation $-12 \mu\text{s}$; $P < 0.001$, paired t -test) towards the ITD value represented at the AGF site (Fig. 1c, arrows).

The opposite result was observed in an experiment in which the best ITDs at the AGF and optic tectum sites differed by $45 \mu\text{s}$ ($\sim 18^\circ$ azimuth; Fig. 2). AGF microstimulation (best ITD $-10 \mu\text{s}$; $7 \mu\text{A}$) suppressed auditory responses, but only when the ITD of the sound was near the best ITD for the optic tectum site (best ITD $-55 \mu\text{s}$).

We tested the effects of microstimulation (mean current $14 \mu\text{A}$, range $5\text{--}40 \mu\text{A}$) at 41 AGF sites on auditory responses at 95 optic tectum sites. At 55 sites in the optic tectum (8 single units) for which $\Delta\text{ITD}_{\text{AGF-OT}} \leq 15 \mu\text{s}$ ($\leq 6^\circ$ azimuth), AGF activation either enhanced auditory responses ($n = 45/55$ sites; $P < 0.05$, paired t -test) or had no effect ($n = 10/55$ sites) (Fig. 2e). Because sites with such similar $\Delta\text{ITD}_{\text{AGF-OT}}$ showed a similar effect, we refer to these collectively as ‘aligned pairs’. At 40 sites in the optic tectum (9 single units) for which $\Delta\text{ITD}_{\text{AGF-OT}} > 15 \mu\text{s}$, AGF activation either suppressed auditory responses ($n = 23/40$ sites; $P < 0.05$, paired t -test) or had no effect ($n = 17/40$ sites) (Fig. 2e); we refer to these as ‘non-aligned pairs’. Single unit and multi-unit sites produced similar results.

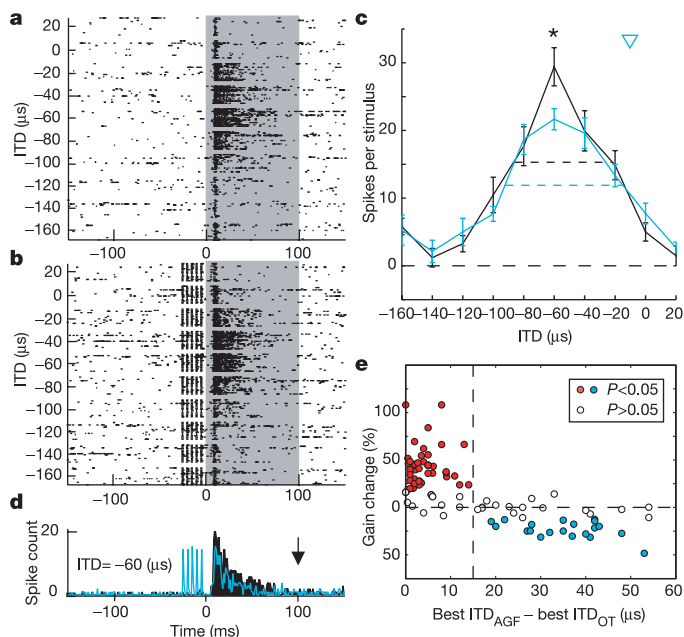


Figure 2 | Effect of AGF microstimulation on a non-aligned site in the optic tectum. The AGF site is the same as in Fig. 1 (best ITD $-10 \mu\text{s}$). For **a–d**, the conventions are as described in Fig. 1. **a**, Raster display for trials without AGF microstimulation. Best ITD $-55 \mu\text{s}$, tuning width $64 \mu\text{s}$. **b**, Raster display for trials with AGF stimulation ($7 \mu\text{A}$, 25-ms train, 200 Hz; current to evoke eye saccades $175 \mu\text{A}$). Best ITD $-55 \mu\text{s}$, tuning width $77 \mu\text{s}$. **c**, Mean responses as a function of ITD for trials with (blue) and without (black) AGF stimulation. Error bars represent s.e.m. Asterisk, $P < 0.03$ (paired t -test). Open arrowhead shows best ITD at the AGF site. **d**, Peristimulus time histograms of optic tectum responses to ITD $-60 \mu\text{s}$ with (blue) and without (black) AGF microstimulation. Arrow shows offset of auditory stimulus. **e**, The effect of AGF microstimulation on optic tectum responses depends on the alignment of AGF and optic tectum tuning. Ordinate shows gain change at each optic tectum site; abscissa shows the difference between ITD tuning of the AGF and optic tectum sites. Filled symbols, $P < 0.05$; open symbols, $P > 0.05$ with current levels up to $40 \mu\text{A}$. Vertical dashed line indicates best $\text{ITD}_{\text{AGF}} - \text{best ITD}_{\text{OT}} = 15 \mu\text{s}$.

The effect of AGF stimulation on optic tectum responses was summarized by weighted averages of the data from individual sites. Across all 55 aligned pairs of sites (Fig. 3a), responses to best ITD increased by an average of 33% (range -8 to 105% , $P < 0.0001$, paired t -test) and the tuning width at half-max decreased by 16% (range -48 to 20% , $P < 0.001$, paired t -test). AGF microstimulation also tended to shift optic tectum tuning at aligned sites towards the value represented at the AGF stimulation site (Fig. 3b).

In contrast, across all 40 non-aligned pairs of sites (Fig. 3c), responses to best ITD decreased by an average of 16% (range -48 to 14% ; $P < 0.0001$, paired t -test) and tuning width at half-max did not change significantly (range -14% to 19% , $P = 0.065$, paired t -test).

The sharpening of auditory tuning curves that resulted from AGF activation is remarkable, because auditory tuning curves measured in the optic tectum under normal conditions are the sharpest of any observed in any part of the brain in any species (average width at half-max $40 \mu\text{s}$, or 16° azimuth)^{22,24}. Our results show that with AGF activation, they become even sharper (average width at half-max $32 \mu\text{s}$, or 12° azimuth). This sharpening is caused by an enhancement of responses specifically to ITDs within $10 \mu\text{s}$ of the best ITD. This modulation of responsiveness that operates selectively within a site’s receptive field cannot be accounted for by non-selective gain changes occurring at a stage in the pathway where receptive fields are smaller, as has been suggested to explain analogous effects in extrastriate visual areas^{25,26}, because auditory receptive fields at earlier stages in

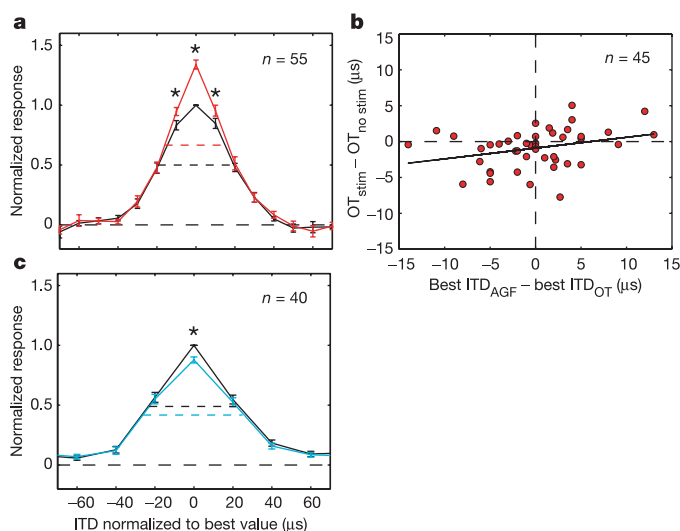


Figure 3 | Summary of the effects of AGF microstimulation on auditory responses in the optic tectum. **a**, Population ITD tuning curves for aligned pairs ($n = 55$) measured with (red) and without (black) AGF microstimulation. Population tuning curves were constructed by normalizing the mean response at each optic tectum site to the maximum mean response for that site without AGF stimulation. All tuning curves were aligned according to their best value (plotted as ITD 0 μ s) and averaged. Error bars represent s.e.m. Note that here, AGF best ITDs were to either side of optic tectum best ITDs. Dashed lines indicate width at half-max. Asterisk, $P < 0.01$ (paired t -test). **b**, Shift in optic tectum tuning induced by AGF microstimulation ($OT_{stim} - OT_{nostim}$) was positively correlated with the difference in ITD tuning between the AGF and the optic tectum (Best ITD_{AGF} - best ITD_{OT}) ($n = 45/55$ aligned pairs, Pearson correlation coefficient $r = 0.3$, $P < 0.05$, two-tailed Student's t -test). **c**, Same as **a**, except that the tuning curves are based on data collected from non-aligned pairs ($n = 40$). Error bars represent s.e.m. Asterisk, $P < 0.01$ (paired t -test).

this pathway are larger. Thus, this result indicates that a gain control mechanism with extremely high spatial resolution can increase auditory response gain within a portion of the site's receptive field.

AGF activation causes an increase in response gain for the ITDs represented by the AGF site (Fig. 3b). The same AGF activation causes a suppression of responses to auditory stimuli at other locations in space (Fig. 2). The result is a top-down enhancement of the representation of auditory stimuli, selected on the basis of stimulus location by this gaze control circuitry. These effects are similar not only to the microstimulation-induced modulations of visual responses observed in monkeys⁹ (effects that have been linked directly to spatial attention¹⁰) but also to modulations of visual responses observed in animals trained to direct spatial attention^{27–29}. Although behavioural experiments will be required to test whether our results in owls are linked to spatial attention, the similarity of the effects of microstimulating forebrain gaze control circuitry in barn owls (on auditory processing) and in monkeys (on visual processing) suggests a common strategy used by the central nervous system for top-down, space-specific control of sensory gain and offers a path for discovering the underlying mechanisms.

METHODS

Animal preparation and recording. Twelve barn owls were used for this study. Owls were housed in large communal aviaries. For all surgical procedures, owls were anaesthetized with halothane (1%) mixed with nitrous oxide and oxygen (45:55 ratio). General surgical and experimental procedures, described in detail previously³⁰, were approved by the Animal Care and Use Committee of Stanford University and were in accordance with National Institutes of Health and Society for Neuroscience guidelines.

Microstimulation. Electrical microstimulation was delivered with a Grass stimulator (S88) and two Grass stimulus isolation units (PSIU-6). Current amplitude was monitored by way of the voltage drop across a 1-k Ω resistor in

series with the return path of the current source. Electrical stimulation was delivered to the AGF site through an epoxy-coated tungsten microelectrode (0.5–1.0 M Ω impedance measured at 1 kHz). In each bird, the AGF was localized on the basis of protocols and coordinates described previously²³. During each experimental session, the current threshold to evoke a motor response was determined by incrementally increasing the stimulation current until a small-amplitude eye movement (a small deflection in the position of a retinal landmark, pecten oculus, viewed ophthalmoscopically) was observed. Once the threshold for eliciting a motor response from the AGF site was determined, the amplitude of the experimental current pulses was then set to low levels (5 μ A) and incrementally increased until either we observed an effect on optic tectum auditory responses or the current amplitude reached 40 μ A (lowest observed motor threshold 55 μ A).

Auditory stimulation. Auditory tuning was measured by presenting noise bursts (100-ms duration, 4–12 kHz, 0-ms rise/fall time, 20 dB above unit threshold) dichotically as described previously^{22,30}. Tuning for ITD was assessed by presenting 10–20 series of noise bursts with ITD varied in a random, interleaved fashion while ILD was held at the optimal value for the site.

Data analysis. Unit recordings were collected from 12 owls (number of unit recordings per owl: 8, 5, 6, 10, 6, 8, 9, 6, 9, 8, 11, 9). Net responses at each optic tectum site were quantified by subtracting the number of spikes that occurred during the 100-ms interval before stimulus presentation (baseline activity) from the number of spikes occurring during the 100 ms after stimulus (sound) onset. On trials during which microstimulation was applied to the AGF, the stimulation artefact was excluded from the analysis and the baseline spike count was normalized to a 100-ms window before the subtraction. Net responses from each trial were then averaged. Paired t -tests were used to compare net responses during trials with and without stimulation. The significance of dynamic shifts in the best ITDs of optic tectum sites was tested by comparing 15 independent measurements of best ITD, with and without AGF microstimulation, with a two-tailed paired t -test.

Received 7 October; accepted 7 November 2005.

- Moore, T., Armstrong, K. M. & Fallah, M. Visuomotor origins of covert spatial attention. *Neuron* **40**, 671–683 (2003).
- Rizzolatti, G., Riggio, L., Dascola, I. & Umiltà, C. Reorienting attention across the horizontal and vertical meridians: evidence in favor of a premotor theory of attention. *Neuropsychologia* **25**, 31–40 (1987).
- Hoffman, J. E. & Subramaniam, B. The role of visual attention in saccadic eye movements. *Percept. Psychophys.* **57**, 787–795 (1995).
- Shepherd, M., Findlay, J. M. & Hockey, R. J. The relationship between eye movements and spatial attention. *Q. J. Exp. Psychol. A* **38**, 475–491 (1986).
- Corbetta, M. Frontoparietal cortical networks for directing attention and the eye to visual locations: identical, independent, or overlapping neural systems? *Proc. Natl Acad. Sci. USA* **95**, 831–838 (1998).
- Moore, T., Tolia, A. S. & Schiller, P. H. Visual representations during saccadic eye movements. *Proc. Natl Acad. Sci. USA* **95**, 8981–8984 (1998).
- Fischer, B. & Boch, R. Enhanced activation of neurons in prelunate cortex before visually guided saccades of trained rhesus monkeys. *Exp. Brain Res.* **44**, 129–137 (1981).
- Muller, J. R., Philiastides, M. G. & Newsome, W. T. Microstimulation of the superior colliculus focuses attention without moving the eyes. *Proc. Natl Acad. Sci. USA* **102**, 524–529 (2005).
- Moore, T. & Armstrong, K. M. Selective gating of visual signals by microstimulation of frontal cortex. *Nature* **421**, 370–373 (2003).
- Moore, T. & Fallah, M. Microstimulation of the frontal eye field and its effects on covert spatial attention. *J. Neurophysiol.* **91**, 152–162 (2004).
- Cavanaugh, J. & Wurtz, R. H. Subcortical modulation of attention counters change blindness. *J. Neurosci.* **24**, 11236–11243 (2004).
- Rorden, C. & Driver, J. Does auditory attention shift in the direction of an upcoming saccade? *Neuropsychologia* **37**, 357–377 (1999).
- Hikosaka, O. & Wurtz, R. H. Visual and oculomotor functions of monkey substantia nigra pars reticulata. I. Relation of visual and auditory responses to saccades. *J. Neurophysiol.* **49**, 1230–1253 (1983).
- Hikosaka, O., Sakamoto, M. & Usui, S. Functional properties of monkey caudate neurons. II. Visual and auditory responses. *J. Neurophysiol.* **61**, 799–813 (1989).
- Knudsen, E. I. & Knudsen, P. F. Contribution of the forebrain archistriatal gaze fields to auditory orienting behaviour in the barn owl. *Exp. Brain Res.* **108**, 23–32 (1996).
- Knudsen, E. I., Cohen, Y. E. & Masino, T. Characterization of a forebrain gaze field in the archistriatum of the barn owl: microstimulation and anatomical connections. *J. Neurosci.* **15**, 5139–5151 (1995).
- Stanton, G. B., Goldberg, M. E. & Bruce, C. J. Frontal eye field efferents in the macaque monkey: II. Topography of terminal fields in midbrain and pons. *J. Comp. Neurol.* **271**, 493–506 (1988).
- Bruce, C. J., Goldberg, M. E., Bushnell, M. C. & Stanton, G. B. Primate frontal eye fields. II. Physiological and anatomical correlates of electrically evoked eye movements. *J. Neurophysiol.* **54**, 714–734 (1985).

19. Dias, E. C. & Segraves, M. A. Muscimol-induced inactivation of monkey frontal eye field: effects on visually and memory-guided saccades. *J. Neurophysiol.* **81**, 2191–2214 (1999).
20. Knudsen, E. I. & Knudsen, P. F. Disruption of auditory spatial working memory by inactivation of the forebrain archistriatum in barn owls. *Nature* **383**, 428–431 (1996).
21. Knudsen, E. I. Auditory and visual maps of space in the optic tectum of the owl. *J. Neurosci.* **2**, 1177–1194 (1982).
22. Olsen, J. F., Knudsen, E. I. & Esterly, S. D. Neural maps of interaural time and intensity differences in the optic tectum of the barn owl. *J. Neurosci.* **9**, 2591–2605 (1989).
23. Cohen, Y. E. & Knudsen, E. I. Binaural tuning of auditory units in the forebrain archistriatal gaze fields of the barn owl: local organization but no space map. *J. Neurosci.* **15**, 5152–5168 (1995).
24. Harper, N. S. & McAlpine, D. Optimal neural population coding of an auditory spatial cue. *Nature* **430**, 682–686 (2004).
25. Reynolds, J. H. & Chelazzi, L. Attentional modulation of visual processing. *Annu. Rev. Neurosci.* **27**, 611–647 (2004).
26. Motter, B. C. Focal attention produces spatially selective processing in visual cortical areas V1, V2, and V4 in the presence of competing stimuli. *J. Neurophysiol.* **70**, 909–919 (1993).
27. Spitzer, H., Desimone, R. & Moran, J. Increased attention enhances both behavioural and neuronal performance. *Science* **240**, 338–340 (1988).
28. McAdams, C. J. & Maunsell, J. H. Effects of attention on orientation-tuning functions of single neurons in macaque cortical area V4. *J. Neurosci.* **19**, 431–441 (1999).
29. Treue, S. & Maunsell, J. H. Attentional modulation of visual motion processing in cortical areas MT and MST. *Nature* **382**, 539–541 (1996).
30. Miller, G. L. & Knudsen, E. I. Early visual experience shapes the representation of auditory space in the forebrain gaze fields of the barn owl. *J. Neurosci.* **19**, 2326–2336 (1999).

Acknowledgements We thank J. Bergan, M. Cohen, K. Maczko, T. Moore and I. Witten for reviews of earlier versions of the manuscript, and P. Knudsen for technical assistance. This work was supported by grants from the National Institutes of Health (E.I.K.) and an NRSA postdoctoral fellowship (D.E.W.)

Author Contributions D.E.W. and E.I.K. designed the experiment and co-wrote the paper. D.E.W. carried out the electrophysiological recordings and data analysis.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to D.E.W. (dwinkows@stanford.edu).

LETTERS

Constant darkness is a circadian metabolic signal in mammals

Jianfa Zhang^{1*}, Krista Kaasik^{2*}, Michael R. Blackburn¹ & Cheng Chi Lee¹

Environmental light is the 'zeitgeber' (time-giver) of circadian behaviour¹. Constant darkness is considered a 'free-running' circadian state. Mammals encounter constant darkness during hibernation². Ablation of the master clock synchronizer, the suprachiasmatic nucleus, abolishes torpor, a hibernation-like state, implicating the circadian clock in this phenomenon^{2,3}. Here we report a mechanism by which constant darkness regulates the gene expression of fat catabolic enzymes in mice. Genes for murine procollipase (*mClps*) and pancreatic lipase-related protein 2 (*mPlrp2*) are activated in a circadian manner in peripheral organs during 12 h dark:12 h dark (DD) but not light-dark (LD) cycles. This mechanism is deregulated in circadian-deficient *mPer1*^{-/-}/*mPer2*^{m/m} mice. We identified circadian-regulated 5'-AMP, which is elevated in the blood of DD mice, as a key mediator of this response. Synthetic 5'-AMP induced torpor and *mClps* expression in LD animals. Torpor induced by metabolic stress was associated with elevated 5'-AMP levels in DD mice. Levels of glucose and non-esterified fatty acid in the blood are reversed in DD and LD mice. Induction of *mClps* expression by 5'-AMP in LD mice was reciprocally linked to blood glucose levels. Our findings uncover a circadian metabolic rhythm in mammals.

Hibernation is an energy conservation mechanism⁴. Unlike a true hibernator, the laboratory mouse can only undergo torpor^{5,6}. During hibernation, an animal departs from LD and enters the DD environment of a den⁷. We proposed that this environmental change is a signal for the initiation of torpor. Microarray studies were used to identify genes that display differential expression in the liver of DD and LD mice (Supplementary Fig. 1). This screen identified a gene encoding CLPS, the enzymatic partner of PLRP2, required for dietary fat degradation^{7,8}. *mClps* expression is restricted to pancreas and the gastrointestinal organs^{7,8}, so its presence in DD mice livers was unexpected. To clarify this observation, we analysed *mClps* expression in liver messenger RNA (mRNA) of wild-type, *mPer1*-null (*mPer1*^{-/-}), *mPer2* mutant (*mPer2*^{m/m}) and circadian-deficient double mutant (*mPer1*^{-/-}/*mPer2*^{m/m}) mice during zeitgeber time (ZT)^{9,10}. Except for three *mPer1*^{-/-}/*mPer2*^{m/m} animals, northern blot analysis showed no detectable *mClps* expression in livers of wild-type, *mPer1*^{-/-} and *mPer2*^{m/m} LD mice (Fig. 1a). By contrast, during circadian time (CT), *mClps* expression was observed in livers from all four genotypes of DD mice (Fig. 1b). Furthermore, *mClps* expression displayed a robust circadian pattern in wild-type but not in *mPer1*^{-/-}, *mPer2*^{m/m} or *mPer1*^{-/-}/*mPer2*^{m/m} DD mice. In addition, the expression of *mClps* was coordinated with that of its enzymatic partner *mPlrp2* in DD mice (Fig. 1a, b). In LD mice, *mClps* expression was found only in pancreas and stomach (Fig. 1c). However, in DD mice, *mClps* expression was observed in all peripheral tissues sampled except brain and kidney (Fig. 1c). The phase of *mClps* expression in peripheral organs of DD mice was similar

(Supplementary Fig. 2a). These observations raised the fundamental question on the biological relevance of this constant-darkness-regulated phenomenon in mammals.

To demonstrate functionality of *mClps* expression in liver, colipase activity assayed with a triacylglycerol substrate ([³H]triolein)¹¹ was observed in liver extracts from DD but not LD mice (Supplementary Fig. 2b). Exposure to light for 5–7 h inhibited both *mClps* and *mPlrp2* expression in liver of DD mice (Supplementary Fig. 2c). Taking these

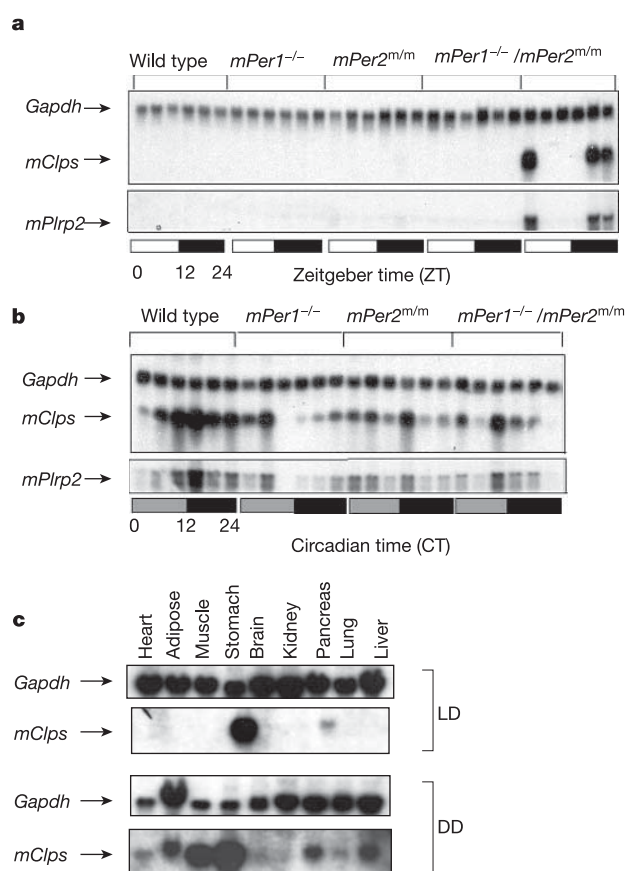


Figure 1 | Northern blot analysis of *mClps* and *mPlrp2* expression in mice livers. **a**, Expression of *mClps* and *mPlrp2* in LD mice. Note: for *mPer1*^{-/-}/*mPer2*^{m/m} samples, the first six lanes from left to right are the corresponding mRNAs from kidney tissues. **b**, Expression of *mClps* and *mPlrp2* in DD mice. **c**, Expression of *mClps* in various peripheral tissues sampled at ZT12 or CT12. *Gapdh* mRNA, encoding glyceraldehyde-3-phosphate dehydrogenase, was monitored as an internal control.

¹Department of Biochemistry and Molecular Biology, University of Texas Health Science Center, Houston, Texas 77030, USA. ²Department of Biotechnology, Institute of Molecular and Cell Biology, Tartu University, Tartu, 51010, Estonia.

*These authors contributed equally to this work.

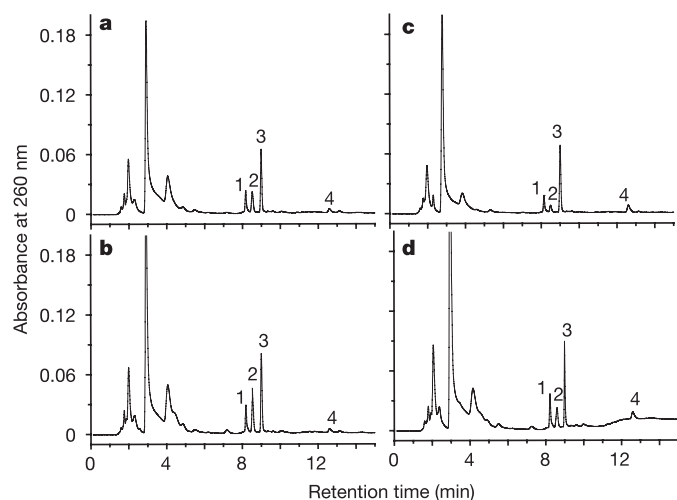


Figure 2 | Elevated concentration of a circadian-regulated circulatory molecule in DD mice. Representative profiles of reverse-phase HPLC analysis of blood extracts taken from LD and DD mice at ZT4 (a), CT4 (b), ZT16 (c) and CT16 (d). Void volume peaks with a retention time of less than 5 min are poorly resolved.

results together, we proposed that *mClps* expression in DD mice is mediated by a circulatory factor that functions either as a repressor or an activator during the LD or DD cycles, respectively. Such an activator would induce *mClps* expression in LD mice but a repressor would inhibit its expression in DD animals. To identify the putative circulatory mediator, blood extracts obtained from mice at various ZTs and CTs were fractionated by HPLC. Excluding the unresolved peaks in the void volume, there were four highly reproducible peaks (labelled 1–4). One peak (no. 2) had a robust apparent diurnal and circadian pattern in both ZT and CT samplings (Fig. 2). Our analysis indicated that peaks 1 and 3 had no apparent diurnal pattern but that peak 4 might have had a weak apparent circadian profile (Supplementary Fig. 3a). A paired *t*-test analysis revealed that only peak 2 was substantially higher in DD mice than in LD mice ($n = 4$, $P < 0.01$; Supplementary Fig. 3b). Spectral scanning of peak 2 revealed a maximum absorbance at 260 nm, suggesting a nucleotide-based molecule. The retention times of peaks 2 and 4 on HPLC matched those of 5'-AMP and adenosine, respectively (Supplementary Fig. 4). The identity of peak 2 was confirmed with snake-venom 5'-nucleotidase, which degrades 5'-AMP (peak 2) to adenosine (peak 4) (Supplementary Fig. 4).

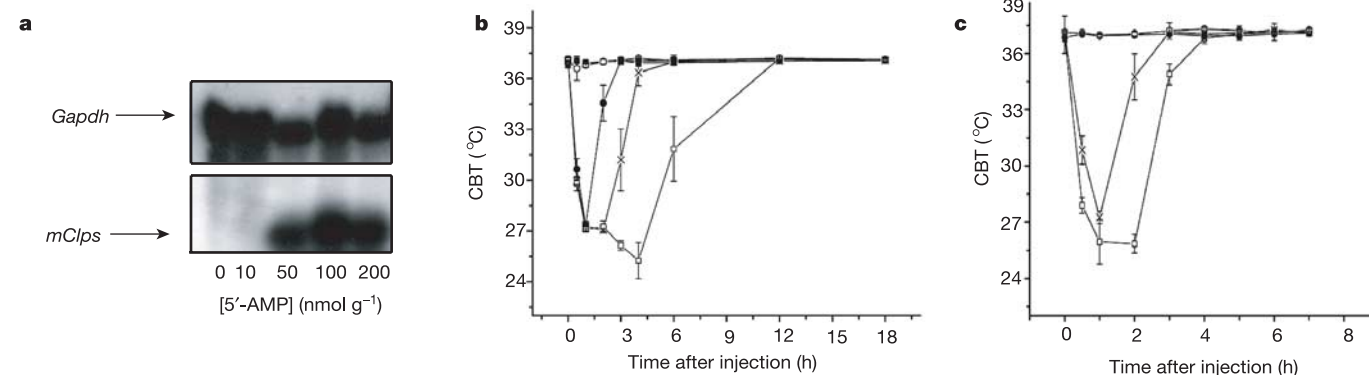


Figure 3 | 5'-AMP-induced *mClps* expression and torpor in LD mice. a, Northern blot analysis of *mClps* expression in liver of wild-type mice injected with saline or 5'-AMP. *Gapdh* levels were monitored as an internal control. b, CBT of wild-type mice injected with saline or 5'-AMP. Filled squares, saline; open circles, 0.15 $\mu\text{mol AMP g}^{-1}$; filled circles, 1.5 $\mu\text{mol AMP g}^{-1}$; crosses, 5.0 $\mu\text{mol AMP g}^{-1}$; open squares, 10.0 $\mu\text{mol AMP g}^{-1}$.

To confirm 5'-AMP as the circulatory factor, we injected synthetic 5'-AMP into LD mice to test the induction of *mClps* expression. Northern blot analysis showed that 5'-AMP induced *mClps* expression in the livers of LD mice at 3.5–4 h after injection (Fig. 3a, and Supplementary Fig. 8b). Using reverse transcriptase-mediated polymerase chain reaction (RT-PCR), we could detect the induction of *mClps* expression by 5'-AMP in all peripheral tissues sampled except brain (Supplementary Fig. 5a). Ecto-5'-nucleotidase anchored on the plasma membrane converts 5'-AMP to adenosine extracellularly^{12,13}. Adenosine receptors or nucleoside transporters could therefore mediate the intracellular action of 5'-AMP. Adenosine but not *N*-ethylcarboxamidoadenosine (NECA), an adenosine receptor agonist, injected into LD mice induced *mClps* expression in liver (Supplementary Fig. 5b, and data not shown). Dipyridamole, a nucleoside transporter blocker¹⁴, prevented *mClps* induction by adenosine and 5'-AMP (Supplementary Fig. 5c). Mice injected with ATP, ADP or c-AMP at similar concentrations did not induce *mClps* expression in liver (Supplementary Fig. 5d). Unexpectedly, LD mice given a high dosage of 5'-AMP had a lower body temperature, suggesting that the animals were in torpor. Mice are in torpor when the core body temperature (CBT) decreases to 31 °C or below^{5,6}. On the basis of CBT measurement, torpor duration in LD mice was dependent on the dosage of 5'-AMP injected (Fig. 3b). Torpor induced by 5'-AMP was significantly longer in *mPer1*^{-/-}/*mPer2*^{m/m} mice than in wild-type animals (Fig. 3c). Together, these studies show that 5'-AMP is the circadian signal that mediates *mClps* expression in peripheral organs and induces torpor in mice.

A question arising from these observations is the biological purpose of this signalling mechanism. Perhaps this circadian signalling mechanism has a function in energy conservation. Therefore, we compared the behaviour of DD mice fed *ad libitum* with that of mice subjected to metabolic stress generated by food deprivation. CBT sampled every 4 h revealed that all of the fasted mice displayed spontaneous torpor by day 2, whereas the CBT of fed mice remained at 37 °C (Fig. 4a). HPLC analysis revealed that 5'-AMP levels in the blood of torpid mice were elevated compared with those of non-torpid DD animals (*t*-test $P < 0.05$; Fig. 4b, c). Thus, under metabolic stress, physiological control of 5'-AMP levels induces torpor in DD mice.

Cessation of food intake and the generation of energy from fat catabolism are hallmarks of deep torpor. The activation of *mClps* expression by constant darkness is probably physiological because murine *mClps* mRNA encodes a pentapeptide (VPDPR) that is cleaved post-translationally from the procolipase enzyme. This pentapeptide is the satiety regulator, enterostatin¹⁵. The DD mice

AMP g^{-1} . Error bars indicate s.e.m. ($n = 3$). c, CBT of wild-type and *mPer1*^{-/-}/*mPer2*^{m/m} mice injected with saline or 5'-AMP. *mPer1*^{-/-}/*mPer2*^{m/m} mice: open squares, 1.5 $\mu\text{mol AMP g}^{-1}$; filled circles, saline. Wild-type mice: crosses, 1.5 $\mu\text{mol AMP g}^{-1}$; open circles, saline. Error bars indicate s.e.m. ($n = 3$). We observed no apparent adverse effects on the torpid mice after their CBT had returned to 37 °C.

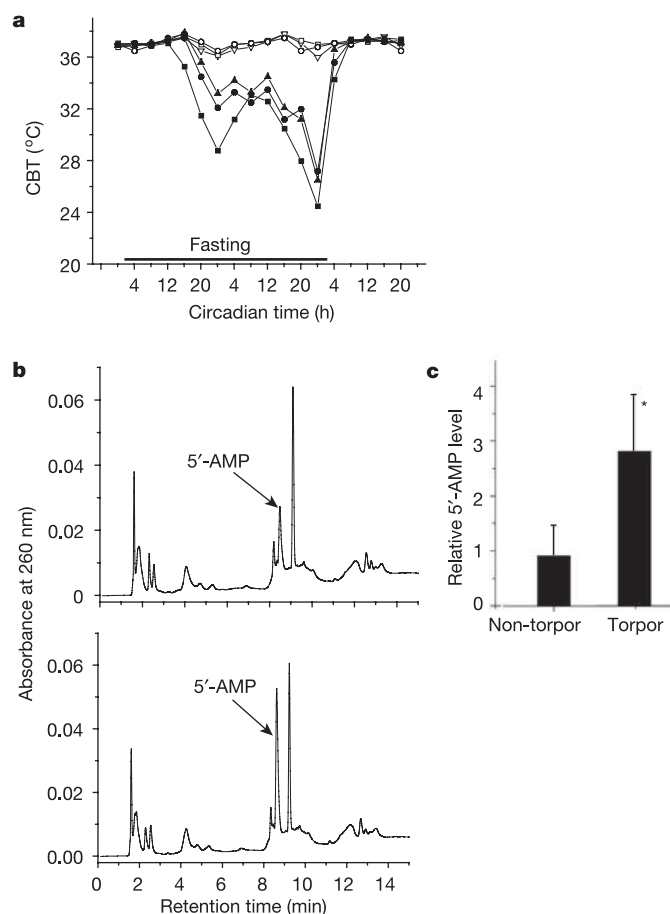


Figure 4 | Torpor and blood 5'-AMP levels in DD mice under metabolic stress. **a**, CBT of fed (open symbols) and fasted (filled symbols) DD mice at ambient room temperature (23 °C). **b**, Representative HPLC analysis of blood extracts from a non-torpid DD mouse (top) and a torpid DD mouse (bottom). **c**, Relative levels of 5'-AMP in torpid and non-torpid DD mice. The average value of 5'-AMP levels from non-torpid mice is arbitrarily set at 1. Error bars indicate s.e.m. ($n = 3$). Asterisk, $P < 0.05$ (paired t -test).

consumed less food and water than the LD animals (Supplementary Fig. 6a and 6b), which is consistent with previous studies on rats¹⁶. Correspondingly, the body weight of DD mice declined during the period studied (Supplementary Fig. 6c). Our studies showed that blood levels of non-esterified fatty acids of DD mice were higher than those in LD animals (Supplementary Fig. 6d), which is consistent with previous observations of large mammals in DD or during denning^{17,18}. Together, these studies demonstrate that the induction of *mClps* expression by constant darkness accomplishes both satiety reduction and the activation of fat catabolism.

Membrane-anchored and circadian-regulated ecto-5'-nucleotidase controls the extracellular level and mediates the intracellular action of 5'-AMP^{12,13,19,20}. Northern blot analysis confirmed that expression of the ecto-5'-nucleotidase gene in LD mice is regulated in a circadian manner and is dampened in DD animals (Supplementary Fig. 7). Ecto-5'-nucleotidase dephosphorylates 5'-AMP to adenosine, which is taken into the cell by nucleoside transporters. Intracellular adenosine is primarily phosphorylated to 5'-AMP by adenosine kinase because its K_m for adenosine is one or two orders of magnitude lower than that of adenosine deaminase¹⁹. Four key metabolic enzymes are regulated allosterically by 5'-AMP. One of these, AMP-dependent protein kinase (AMPK), is activated by 5'-AMP²¹. A 5'-AMP analogue, 5-aminoimidazole-4-carboxamide ribonucleoside (AICAR), increases fatty acid oxidation in rat muscle, presumably through AMPK²². Another enzyme, fructose-1,6-diphosphatase (FDP), is negatively regulated by 5'-AMP²³. FDP,

the rate-limiting enzyme in gluconeogenesis, converts fructose 1,6-diphosphate to fructose 6-phosphate. A third allosteric enzyme, phosphofructokinase (PFK), converts fructose 6-phosphate into fructose 1,6-diphosphate and is positively regulated by 5'-AMP²³. PFK is a rate-limiting enzyme for glycolysis. Consistent with previous studies in rats²⁴ was the observation that blood glucose was lower in DD than in LD mice (Supplementary Fig. 6e). Furthermore, activation of *mClps* expression by 5'-AMP in LD mice is reciprocally related to blood glucose levels (Supplementary Fig. 8a, b). We propose that when 5'-AMP was injected into mice, the activity of FDP was inhibited and that of PFK was enhanced. Consequently, the rate of gluconeogenesis was reduced but that of glycolysis was enhanced, leading to depletion of the blood glucose pool. The transient rise in blood glucose concentration is a first-level metabolic response to replenish this pool. The rate-limiting enzyme glycogen phosphorylase, which converts stored glycogen into glucose 1-phosphate, is another 5'-AMP-activated allosteric enzyme²⁵. When depletion of stored glycogen reaches a critical stage, blood glucose levels decline. To conserve glucose necessary for brain function (see Fig. 1c and Supplementary Fig. 5a), the primary energy source of peripheral organs was switched from glucose to fatty acids, as reflected by induction of *mClps* expression (Supplementary Fig. 8b). Hence, 5'-AMP is a pivotal metabolic signal whose circulatory level determines the balance of the peripheral organ energy supply between glucose, glycogen and fat (Supplementary Fig. 8c). Consistent with this proposition, 5'-AMP does not activate *mClps* expression in cultured cells whose primary energy source is glucose. If such a mechanism is conserved in humans, the action of 5'-AMP and its analogues could form a new class of therapeutic agents for human obesity and insulin-resistant type-2 diabetes. The ability of 5'-AMP to induce torpor could be a useful tool in CBT management during major surgery or emergency trauma response.

Last, a quirky enigma of biochemistry is the 'futile cycle' burning up ATP molecules between FDP and PFK activities²³. Because the endogenous clock controls 5'-AMP levels, the 'futile cycle' is a circadian metabolic rhythm.

METHODS

Animals. We used female mice aged between 8 and 10 weeks. Wild-type (C57/Bl6), *mPer1*^{-/-}, *mPer2*^{m/m} and *mPer1*^{-/-}/*mPer2*^{m/m} mice were housed in a standard animal maintenance facility under a 12h light:12h dark cycle^{9,10}. For 12h dark:12h dark (DD) studies, mice were placed inside a circadian chamber beginning at CT12 for 48 h under constant darkness before the mice were used for the indicated experiments. All manipulations of DD mice were performed under a 15-W red light²⁶. These studies were conducted under institutionally approved animal protocol HSC-AWC 04-022.

Northern blot and RT-PCR analysis. Tissues were collected and frozen in liquid nitrogen and stored at -80 °C. Total RNA was isolated from mouse livers in accordance with standard procedures²⁷. Northern blot analysis was performed as described previously²⁶. The colipase probe was the complete complementary DNA (GenBank accession no. BC042935); the *Gapdh* probe was the *Pst*I fragment of rat *Gapdh* cDNA²⁸. The primer pair used to measure colipase expression was 5'-TTGTTCTTCTGCTTGTCCT-3' and 5'-AGTCGAGGCAGATGCCATAGTT-3'. The primer pair used to measure *Gapdh* expression as an internal control was 5'-AAGCCCATCACCATTCTTCCA-3' and 5'-ATGGCATGGACTGTGGTCAT-3'. A 720-base-pair probe for *mPlrp2* was generated by RT-PCR with oligonucleotides LipaseF (5'-CGGTGGACCATCGGATGCATG-3') and LipaseR (5'-GAAGTCTTTCCCGTCTTTACCGCG-3') from liver mRNA.

Hepatic colipase activity assay. Livers were removed from mice under ambient light (ZT0 and ZT12) or under a 15-W red light (CT0 and CT12) and protein extracts were prepared as described previously⁸. The samples were heated for 15 min at 65 °C to inactivate endogenous lipases. The protein content of the extracts was determined by the bicinchoninic acid method (Pierce). The heat-inactivated samples were assayed for the presence of colipase with [³H]triolein as substrate, as described previously¹¹.

HPLC analysis of adenine nucleotides. Blood was rapidly removed from mice and frozen in liquid nitrogen. Nucleotides were extracted from frozen samples with 0.4 M perchloric acid as described previously²⁹. Blood extracts and adenine nucleotides ATP, ADP, AMP, c-AMP and adenosine (Sigma) were separated and

quantified by reverse-phase HPLC (Waters, Millipore Corp.) analysis on a Partisphere-bonded phase C₁₈ (reverse-phase) cartridge column at a flow rate of 1.5 ml min⁻¹ (ref. 29). The mobile phase was 0.02 M NH₄H₂PO₄ pH 5.1 with a superimposed methanol gradient with the following time course: 0% for 0–4 min, 0–8% for 4–6 min, 8–20% for 6–8 min and 20% for 8–18 min.

Injection of 5'-AMP, adenosine, NECA and dipyridamole. The indicated dosages of 5'-AMP, adenosine, NECA and dipyridamole (Sigma) were administered to LD mice by intraperitoneal injection. NECA was administered at 0.3 nmol g⁻¹ body weight. All injections took place at ZT6. After injection, mice were maintained for the desired duration (2.5–3.0 h for adenosine or NECA, and 3.5 h for 5'-AMP) and then killed. Total RNA was isolated from liver tissue for northern blot and RT-PCR analysis. Core body temperature (CBT) was measured at ambient room temperature (23–24 °C) with a rectal thermometer before and after each injection.

Metabolic stress studies. Core body temperature and 5'-AMP levels in blood during the fasting were measured in fed or fasted DD mice. The fasted DD mice had their chow removed starting at CT2. Torpor was detected by CBT measurement, and animals in torpor were either killed for blood samples or given food at the third CT2. Food and water intakes were determined by weight differential of fresh chow and water after every 24 h at ZT2 or CT2. Body weight was measured at every ZT2 or CT2. Glucose and non-esterified fatty acid levels in serum were measured with a glucose assay kit from BioAssay Systems and a non-esterified fatty acid assay kit from Roche Applied Science, respectively.

Received 1 July; accepted 26 October 2005.

- Aschoff, J. Exogenous and endogenous components in circadian rhythms. *Cold Spring Harb. Symp. Quant. Biol.* **25**, 11–28 (1960).
- Ruby, N. F., Dark, J., Heller, H. C. & Zucker, I. Ablation of suprachiasmatic nucleus alters timing of hibernation in ground squirrels. *Proc. Natl Acad. Sci. USA* **93**, 9864–9868 (1996).
- Heller, H. C. & Ruby, N. F. Sleep and circadian rhythms in mammalian torpor. *Annu. Rev. Physiol.* **66**, 275–289 (2004).
- Heldmaier, G., Ortman, S. & Elvert, R. Natural hypometabolism during hibernation and daily torpor in mammals. *Respir. Physiol. Neurobiol.* **141**, 317–329 (2004).
- Gavrilova, O. et al. Torpor in mice is induced by both leptin-dependent and -independent mechanisms. *Proc. Natl Acad. Sci. USA* **96**, 14623–14628 (1999).
- Overton, J. M. & Williams, T. D. Behavioral and physiologic responses to caloric restriction in mice. *Physiol. Behav.* **81**, 749–754 (2004).
- Lowe, M. E. Molecular mechanisms of rat and human pancreatic triglyceride lipases. *J. Nutr.* **127**, 549–557 (1997).
- D'Agostino, D. et al. Decreased postnatal survival and altered body weight regulation in prolipase deficient mice. *J. Biol. Chem.* **277**, 7170–7177 (2002).
- Zheng, B. et al. Nonredundant roles of the *mPer1* and *mPer2* genes in the mammalian circadian clock. *Cell* **105**, 683–694 (2001).
- Zheng, B. et al. The *mPer2* gene encodes a functional component of the mammalian circadian clock. *Nature* **400**, 169–173 (1999).
- Lowe, M. E. in *Methods in Molecular Biology: Lipase and Phospholipase Protocols* (eds Doolittle, M. H. & Reue, K.) 59–70 (Humana, Totowa, New Jersey, 1998).
- Thompson, L. F., Ruedi, J. M., Glass, A., Low, M. G. & Lucas, A. H. Antibodies to 5'-nucleotidase (CD73), a glycosyl-phosphatidylinositol-anchored protein, cause human peripheral blood T cells to proliferate. *J. Immunol.* **143**, 1815–1821 (1989).
- Ogata, S., Hayashi, Y., Misumi, Y. & Ikehara, Y. Membrane-anchoring domain of rat liver 5'-nucleotidase: identification of the COOH-terminal serine-523 covalently attached with a glycolipid. *Biochemistry* **29**, 7923–7927 (1990).
- Thorn, J. A. & Jarvis, S. M. Adenosine transporters. *Gen. Pharmacol.* **27**, 613–620 (1996).
- Erlanson-Albertsson, C. & Larsson, A. The activation peptide of pancreatic procolipase decreases food intake in rats. *Regul. Pept.* **22**, 325–331 (1988).
- Stoynev, A. G. & Ikonov, O. C. Effect of constant light and darkness on the circadian rhythms in rats: I. Food and water intake, urine output and electrolyte excretion. *Acta Physiol. Pharmacol. Bulg.* **9**, 58–64 (1983).
- Alila-Johansson, A., Eriksson, L., Soveri, T. & Laakso, M. L. Daily and annual variations of free fatty acid, glycerol and leptin plasma concentrations in goats (*Capra hircus*) under different photoperiods. *Comp. Biochem. Physiol. A Mol. Integr. Physiol.* **138**, 119–131 (2004).
- LeBlanc, P. J. et al. Correlations of plasma lipid metabolites with hibernation and lactation in wild black bears *Ursus americanus*. *J. Comp. Physiol. [B]* **171**, 327–334 (2001).
- Arch, J. R. & Newsholme, E. A. Activities and some properties of 5'-nucleotidase, adenosine kinase and adenosine deaminase in tissues from vertebrates and invertebrates in relation to the control of the concentration and the physiological role of adenosine. *Biochem. J.* **174**, 965–977 (1978).
- von Mayersbach, H. & Klaushofer, K. Circadian variations of 5'-nucleotidase activity in rat liver. *Cell. Mol. Biol. Cyto-enzymol.* **24**, 73–79 (1979).
- Lindsley, J. E. & Rutter, J. Nutrient sensing and metabolic decisions. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* **139**, 543–559 (2004).
- Kaushik, V. K. et al. Regulation of fatty acid oxidation and glucose metabolism in rat soleus muscle: effects of AICAR. *Am. J. Physiol. Endocrinol. Metab.* **281**, 335–340 (2001).
- Lehninger, A. L. *Biochemistry: The Molecular Basis of Cell Structure and Function* 2nd edn 623–657 (Worth, New York, 1977).
- Ahlersova, E., Ahlers, I., Toropila, M. & Smajda, B. Influence of light regimens on circadian changes in the blood glucose and tissue glycogen concentration in the rat. *Physiol. Bohemoslov.* **31**, 57–64 (1982).
- Garcia-Fuentes, L., Camara-Artigas, A., Lopez-Mayorga, O. & Baron, C. Thermodynamic characterization of 5'-AMP binding to bovine liver glycogen phosphorylase a. *J. Biol. Chem.* **271**, 27569–27574 (1996).
- Kaasik, K. & Lee, C. C. Reciprocal regulation of haem biosynthesis and the circadian clock in mammals. *Nature* **430**, 467–471 (2004).
- Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. *Biochemistry* **18**, 5294–5299 (1979).
- Fort, P. et al. Various rat adult tissues express only one major mRNA species from the glyceraldehyde-3-phosphate-dehydrogenase multigenic family. *Nucleic Acids Res.* **13**, 1431–1442 (1985).
- Knudsen, T. B. et al. Effects of (R)-deoxycoformycin (pentostatin) on intrauterine nucleoside catabolism and embryo viability in the pregnant mouse. *Teratology* **45**, 91–103 (1992).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. Lever for helpful comments, and J. Volmer for affinity-purified 5'-nucleotidase and pSK-5' NT. This work was supported in part by an NIH grant and the UTHSC Dean's fund to C.C.L. M.R.B. is supported in part by NIH funding.

Author Contributions J.Z. carried out the described metabolic experiments, characterized peak 2 as 5'-AMP, and demonstrated that 5'-AMP induces torpor and expression of *mClps* in peripheral organs that is blocked by dipyridamole. K.K. screened and identified *mClps/mPlrp2* expression in liver of DD mice. M.R.B. contributed insight into adenosine chemistry. C.C.L. conceived and directed the work and recognized the differential temporal profiles of peak 2 in DD and LD mice.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to C.C.L. (cheng.c.lee@uth.tmc.edu).

LETTERS

Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population

Marco Vignuzzi¹, Jeffrey K. Stone¹, Jamie J. Arnold², Craig E. Cameron² & Raul Andino¹

An RNA virus population does not consist of a single genotype; rather, it is an ensemble of related sequences, termed quasispecies^{1–4}. Quasispecies arise from rapid genomic evolution powered by the high mutation rate of RNA viral replication^{5–8}. Although a high mutation rate is dangerous for a virus because it results in nonviable individuals, it has been hypothesized that high mutation rates create a ‘cloud’ of potentially beneficial mutations at the population level, which afford the viral quasispecies a greater probability to evolve and adapt to new environments and challenges during infection^{4,9–11}. Mathematical models predict that viral quasispecies are not simply a collection of diverse mutants but a group of interactive variants, which together contribute to the characteristics of the population^{4,12}. According to this view, viral populations, rather than individual variants, are the target of evolutionary selection^{4,12}. Here we test this hypothesis by examining the consequences of limiting genomic diversity on viral populations. We find that poliovirus carrying a high-fidelity polymerase replicates at wild-type levels but generates less genomic diversity and is unable to adapt to adverse growth conditions. In infected animals, the reduced viral diversity leads to loss of neurotropism and an attenuated pathogenic phenotype. Notably, using chemical mutagenesis to expand quasispecies diversity of the high-fidelity virus before infection restores neurotropism and pathogenesis. Analysis of viruses isolated from brain provides direct evidence for complementation between members in the quasispecies, indicating that selection indeed occurs at the population level rather than on individual variants. Our study provides direct evidence for a fundamental prediction of the quasispecies theory and establishes a link between mutation rate, population dynamics and pathogenesis.

To examine the biological role of viral quasispecies, we searched for viruses carrying a polymerase with enhanced fidelity, which should decrease genomic diversity and restrict quasispecies complexity. To this end, we isolated poliovirus resistant to ribavirin (Supplementary Fig. S1), a mutagen that increases mutation frequency of poliovirus replication above the tolerable error threshold and drives the virus into viral extinction^{13,14}. The ribavirin-resistant mutant replicated efficiently in the presence of ribavirin, producing over 300-fold more virus than wild type (Supplementary Fig. S2a). The ribavirin-resistant phenotype is determined by a single point mutation, Gly64 to Ser (G64S), in the finger domain of the viral polymerase¹⁵. Notably, the same mutation was independently isolated in another screen for ribavirin-resistant polioviruses¹⁶, suggesting that there are limited mechanistic avenues for overcoming the mutagenic effects of ribavirin.

On the basis of genetic evidence, it was proposed that ribavirin

resistance relies on a ‘super-accurate’, high-fidelity polymerase¹⁶. A lower error rate would reduce the risk of exceeding the tolerable error threshold in response to the mutagen^{13,14}. To examine this possibility, we used direct sequencing of viral isolates within the population to determine whether G64S viral populations generate fewer variants relative to the original genome. Indeed, G64S mutant populations had sixfold fewer mutations than wild-type populations (~0.3 mutations per genome for G64S compared to ~1.9 mutations per genome in wild type) (Table 1), indicating that G64S viral populations are genetically more homogenous. We also assessed population diversity by determining the proportion of a genetic marker present in wild-type and G64S populations. Poliovirus RNA replication is strongly inhibited by the presence of 2 mM guanidine hydrochloride (GdnHCl); however, mutations that confer resistance to GdnHCl (*gua*^r) have been identified^{17,18}. In good agreement with our sequencing data and with previous observations¹⁶, wild-type virus stocks had about 3–4-fold more *gua*^r viruses than the restricted G64S quasispecies (Table 1). Biochemical studies directly confirmed that the G64S RNA polymerase has increased fidelity relative to wild type¹⁹. Although Gly64 is remote from the catalytic centre, this residue participates in a network of hydrogen bonds¹⁵ that influences the conformation of Asp 238, a residue that is critical for nucleotide selection^{19,20}.

The identification of G64S as a mutation that restricts genome diversity in viral populations allowed us to critically examine the quasispecies hypothesis. Given that Gly64 is highly conserved (see <http://www.virology.wisc.edu/acp/Aligns/aligns/entero.p123>), we

Table 1 | Genomic diversity for wild-type, G64S and G64S^{eQS} populations

Virus	Total number of mutations*	Mutations per genome	<i>gua</i> ^r variants†
Wild type	13/50,700	1.91	62 ± 6
G64S	2/50,700	0.31	22 ± 8
G64S ^{eQS}	14/50,700	2.06	56 ± 6
Wild type ^{b,‡}	12/48,588	1.84	ND
G64S ^{b,‡}	6/69,000	0.65	ND
G64S ^{eQS-b,‡}	6/50,700	0.88	ND

*Number of mutations observed over the total number of nucleotides sequenced. To determine the mutation frequency in each poliovirus population, 24 independent poliovirus cDNA clones were obtained. Poliovirus cDNAs were generated by RT-PCR from viral RNA isolated from single plaques in a standard plaque assay. A significant difference in the number of mutations was observed between wild-type and G64S viruses ($P < 0.002$, Mann-Whitney *U*-test). In contrast, no significant difference was observed between wild-type and G64S^{eQS} viruses ($P < 0.222$).

†Per 10⁶ plaque-forming units (p.f.u.), mean ± s.d. of six experiments. Significance testing yielded $P < 0.001$ by analysis of variance (ANOVA).

‡Wild-type^b, G64S^b and G64S^{eQS-b} viruses re-isolated from infected brain. Number of *gua*^r variants was not determined (ND).

¹Department of Microbiology and Immunology, University of California, San Francisco, California 94143-2280, USA. ²Department of Biochemistry and Molecular Biology, Pennsylvania State University, University Park, Pennsylvania 16802, USA.

first examined whether increased fidelity was achieved at the expense of viral replication rate. One-step growth curves and northern blot analysis of genomic RNA synthesis indicate that wild-type and G64S virus replicated at very similar rates and levels (Supplementary Fig. S2b, c). Furthermore, poliovirus replicons carrying the G64S allele replicated with identical efficiency to wild type (Supplementary Fig. S2d). We thus conclude that the G64S mutation confers higher fidelity without significant reduction in the overall RNA replication efficiency.

Greater sequence heterogeneity, characteristic of viruses replicating close to the tolerable error threshold, generates diverse quasi-species, thereby providing a reservoir of mutations that enable virus adaptation to changing environments encountered during infection^{4,11}. To test this postulate, we compared the adaptability of wild-type and G64S-restricted quasiespecies to an experimental environmental stress created by the presence of the poliovirus inhibitor GdnHCl. As expected, the wild-type population adapted at significantly faster rates than G64S to the new environment (Supplementary Fig. S3c). We also examined the evolutionary progression of wild-type and G64S-restricted quasiespecies by determining the spontaneous accumulation of *gua*^r mutants in populations grown under no selective pressure. The higher fidelity polymerase G64S is restricted in its ability to build a reservoir of potentially beneficial mutations (Supplementary Fig. S3a). Furthermore, the increase in polymerase fidelity decreases viral fitness, as defined by the inability of G64S to effectively compete with wild-type virus under adverse growth conditions (Supplementary Fig. S3b). Taken together, these experiments support the idea that the diversity of a quasiespecies is essential for adapting to and surviving new selective pressures in changing environments.

The major challenges to viral survival occur during its interactions with the host. During infection, viruses struggle with host-to-host transmission, host defence mechanisms, diverse cellular environments in different tissues, and anatomical restrictions such as the blood–brain barrier. The outcome of these multiple selective pressures determines tissue tropism and ultimately, the pathogenic

outcome of an infection. To evaluate whether restricting population diversity affects the biological course of a viral infection, G64S was inoculated in susceptible mice by intramuscular injection, which allows poliovirus to quickly access the central nervous system (CNS) by axonal retrograde transport²¹.

G64S virus presented a highly attenuated phenotype, in which onset of paralysis was delayed and observed only at very high viral doses (Fig. 1a). The 50% lethal dose (LD₅₀) for G64S was more than 300-fold higher than for wild type (Fig. 1c). Furthermore, intravenous inoculation indicated that the high-fidelity G64S virus shows restricted tissue tropism. Although both wild-type and G64S virus were readily isolated over several days from the spleen, kidney, muscle and small intestine, G64S virus was unable to establish infection and replicate effectively in the spinal cord and brain (Fig. 2a), despite these being principal sites of wild-type poliovirus replication.

The attenuated phenotype observed for G64S could stem from

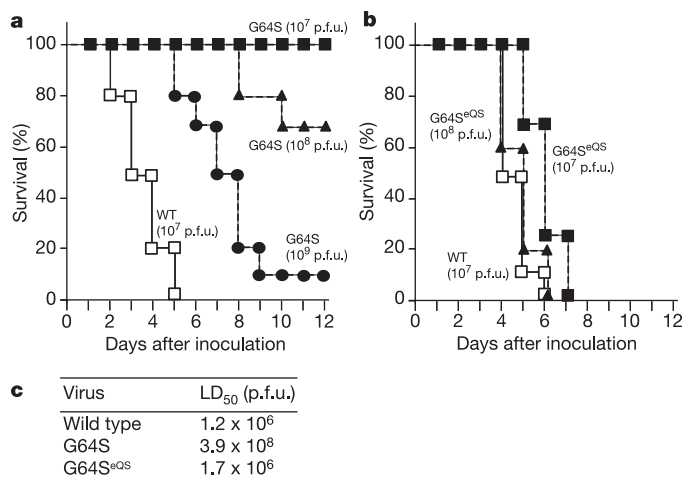


Figure 1 | A restricted quasiespecies of poliovirus is less neuropathogenic. **a, b**, Percentage of mice surviving intramuscular injection of different doses (10⁷, 10⁸ or 10⁹ p.f.u.) of the G64S (**a**) or the expanded G64S^{eQS} (**b**) populations, compared with wild type (WT, open symbols; only one dose (10⁷ p.f.u.) is shown); *n* = 20 mice per group. The differences observed between wild type (10⁷ p.f.u.) and G64S (10⁷, 10⁸ or 10⁹ p.f.u.) and between wild type and the expanded G64S^{eQS} (10⁷ p.f.u.) are statistically significant (*P* < 0.001). In contrast, no statistically significant difference was observed between wild type (10⁷ p.f.u.) and G64S^{eQS} (10⁸ p.f.u.) (*P* > 0.5). **c**, Calculation of LD₅₀ values for each viral stock, using the Reed and Muench method.

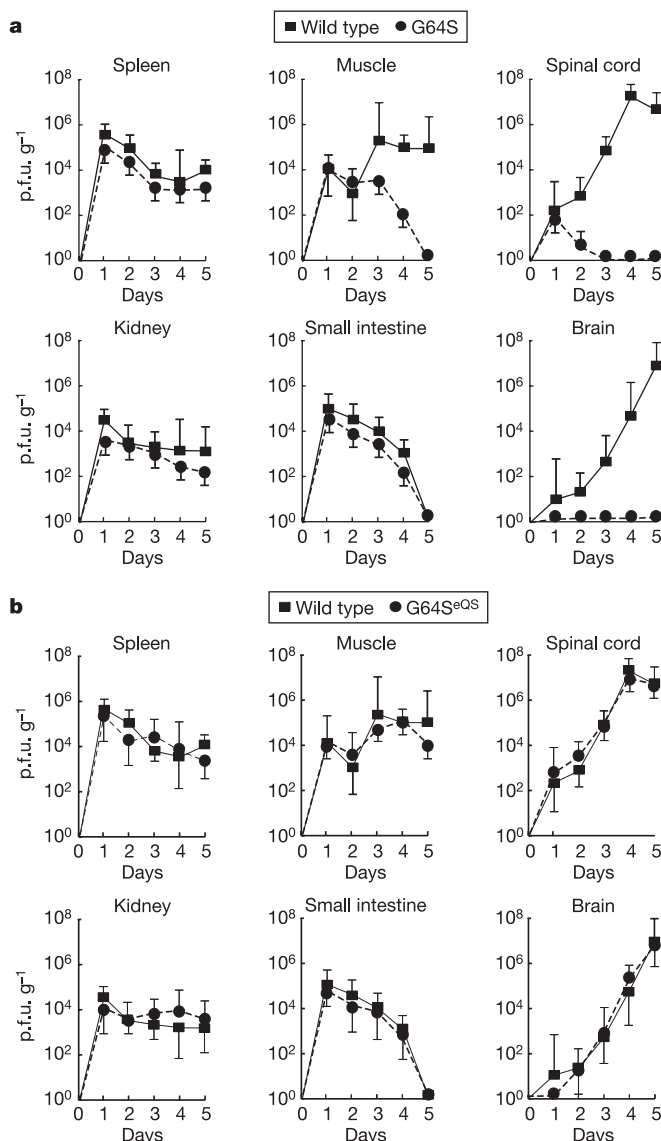


Figure 2 | Genomic diversity in quasiespecies is critical in viral tissue tropism and pathogenesis. **a, b**, Virus titres in p.f.u. per gram from tissue of mice infected intravenously with the wild-type virus population (squares), the narrow G64S quasiespecies (circles in **a**) or the artificially expanded G64S^{eQS} quasiespecies (circles in **b**). Mean values ± s.d. of three experiments are shown.

the restricted nature of its quasispecies diversity or could be the consequence of a specific RNA replication defect caused by the G64S mutation *in vivo*. To distinguish between these possibilities, we created an artificially expanded quasispecies that retains the G64S mutation in the polymerase (Supplementary Fig. S4). Treatment of viral stocks with chemical mutagens (ribavirin and 5-fluorouracil) increased the number of mutations in the G64S genome (G64S^{eQS}, expanded quasispecies) to wild-type levels, as determined by direct sequencing and by the number of *gua*^r viruses present in the population (Table 1). Of more than 25 independent clones sequenced, all viruses in the G64S^{eQS} population conserved the G64S substitution (not shown).

The more diverse G64S^{eQS} viral population showed a significant increase in neuropathogenesis, and the LD₅₀ of the mutagenized population was very similar to wild type (Fig. 1b, c). Furthermore, the tissue distribution of G64S^{eQS} was indistinguishable from that observed for wild type (Fig. 2b). Sequencing of 24 isolates obtained directly from brain tissues of mice infected with G64S^{eQS} confirmed that the G64S substitution was still present in all instances. We thus conclude that the G64S mutation does not in itself preclude replication in neuronal tissues.

We next considered whether the artificial expansion of G64S^{eQS} enhanced pathogenicity by generating specific neurotropic mutations. In the simplest model, viruses carrying these advantageous mutations could selectively enter and replicate in the CNS. Accordingly, we re-isolated G64S^{eQS} from brains of infected mice and analysed them by direct sequencing. Strikingly, the sequence of the viral RNA from brain was indistinguishable from the original inoculated G64S^{eQS} genome, as well as from the genomes of wild-type (apart from the 64 position) or G64S viruses (not shown). Because direct sequencing of the viral RNA population only reveals the predominant consensus sequences of the quasispecies, we also sequenced 72 independently cloned viruses isolated from infected brains. Again, no recurring mutations were detected, suggesting that a discrete set of mutations was not selected in the neurotropic virus population.

As the inability to detect a specific set of mutations by direct sequencing could indicate that a very large set of mutations within the G64S^{eQS} virus is neurotropic, we next used a functional assay to analyse the viral populations isolated from infected brains. If infection of the mouse brain is caused by selection of a neurovirulent set of G64S variants, isolation of this population and subsequent re-inoculation should result in neuropathogenesis. Accordingly, we intravenously re-inoculated virus populations obtained from brains of animals infected with wild-type, G64S or viruses. Strikingly, although wild-type virus remained neurotropic, G64S^{eQS} isolated from brain (G64S^{eQS-b}) was no longer able to infect the CNS (Fig. 3a). This result does not support the idea that a set of neurotropic mutations determines the pathogenic characteristics of G64S^{eQS}. Notably, the G64S^{eQS-b} and G64S^b populations had 2–3-fold fewer mutations than wild-type^b and the original artificially expanded quasispecies G64S^{eQS} (Table 1), indicating that G64S^{eQS-b} lost diversity during *in vivo* replication. Our results are not consistent with the model that expanding the quasispecies repertoire of G64S^{eQS} enhances pathogenesis by generating a defined set of neurotropic mutations. Instead, they suggest a more complex model whereby a generalized increase in sequence diversity determines pathogenesis.

In contrast to classic genetic concepts suggesting that evolution occurs through the selection of individual viruses, the quasispecies theory proposes, on the basis of theoretical considerations, that evolution occurs through selection of interdependent viral subpopulations^{4,9,10}. This alternative model could explain our results. The interplay between different variants within the quasispecies could facilitate entry and replication of the virus population in the CNS. To test this model, we examined the ability of a non-neurotropic virus population carrying an identifier barcode (G64S^{Sac}) to infect the CNS, either by itself or upon co-infection with either of two

neurotropic populations: wild-type virus or the expanded G64S^{eQS}. G64S^{Sac} carries the G64S allele that restricts genome diversity, as well as a neutral *SacI* restriction site that can be used to identify its RNA (Fig. 3b and Supplementary Fig. S6).

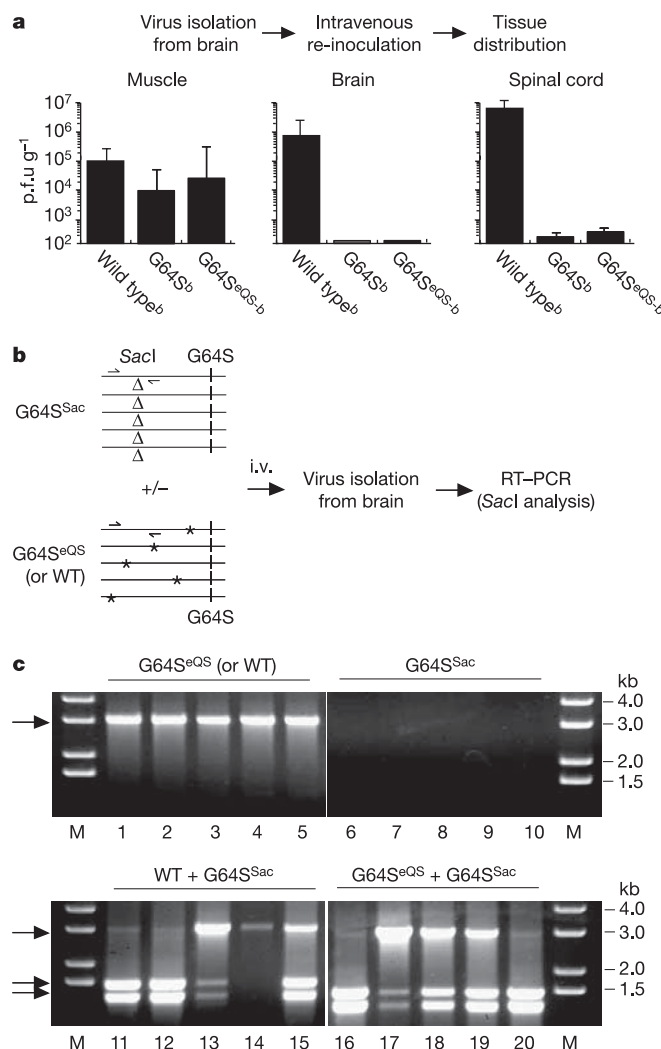


Figure 3 | Cooperative interactions among members of the virus population link quasispecies diversity with pathogenesis. **a**, Subpopulations of viruses isolated from brains of infected mice cannot re-establish CNS infection if the diversity of the quasispecies is restricted. Virus titres (p.f.u. g⁻¹) from muscle, brain and spinal cord of mice infected intravenously for 4 days with 10⁷ p.f.u. viruses isolated from brains of infected animals with wild type^b, G64S^b, G64S^{eQS-b}. A schematic representation of the re-inoculation protocol is also shown. **b**, **c**, Neurotropic virus populations facilitate entry and replication of a non-neurotropic virus into the CNS. **b**, Schematic representation of an *in vivo* complementation experiment. G64S^{Sac} is a narrow quasispecies virus carrying a higher fidelity polymerase (G64S allele) and a silent mutation that introduces a *SacI* site at nucleotide 1906 within the capsid region. This neutral genetic marker can be used as a 'barcode'. Mice were infected intravenously with either G64S^{Sac} alone (2 × 10⁸ p.f.u. per animal) or co-injected with wild type (WT) or G64S^{eQS} at 1:1 ratios (10⁸ p.f.u. of each virus per animal). Viruses were re-isolated from brain tissues, through infection of HeLa cells, and their RNA was analysed by RT-PCR. **c**, All PCR products were digested with *SacI* before analysis by agarose gel electrophoresis. DNA obtained from wild-type or G64S^{eQS} viruses were not digested by *SacI* (~3 kb fragment); whereas the PCR product from the G64S^{Sac} virus generated two smaller bands (~1.55 and 1.45 kb) when digested with *SacI*. The injected viruses are indicated at the top of the gel. Each lane corresponds to one infected mouse. Arrows on the left indicate full-length RT-PCR products and *SacI*-digested PCR fragments. DNA markers (kb) are shown in lane M.

After intravenous inoculation, viruses were isolated from the brain and analysed by polymerase chain reaction with reverse transcription (RT–PCR) and *SacI* digestion. As expected, although wild-type or G64S^{eqs} viruses alone were readily isolated from brain, G64S^{sac} was unable to infect either brain (Fig. 3c, upper panel and Supplementary Fig. S6b) or spinal cord (not shown). In contrast, G64S^{sac} virus was isolated from the brain tissue of every infected mouse if co-inoculated with either wild-type or G64S^{eqs} virus (Fig. 3c, lower panel). To confirm this observation, we cloned and analysed individual viruses from brain homogenates obtained from infected mice. In co-infections, approximately 50% of the viruses isolated from brain corresponded to G64S^{sac} (Supplementary Table S1). These data indicate that there is a positive interaction between different variants within the population, with some variants allowing others to enter the brain—as anticipated by the quasispecies theory.

Our study shows that increasing the fidelity of poliovirus replication has a marked effect on viral adaptation and pathogenicity. These findings, alongside previous observations that an increase in error rate above the tolerable error threshold leads to viral extinction^{22–24}, suggest that the viral mutation rate is finely tuned and has probably been optimized during evolution of the virus. Survival of a given viral population depends on the balance between replication fidelity, which ensures the transmission of its genetic makeup, and genomic flexibility, which allows build up of a reservoir of individual variants within the population that facilitates adaptation to changing environmental conditions. Although having too many mutations in a genome can drive a viral population to extinction^{13,14}, too few mutations can cause extinction by rendering the virus unable to survive changes in the environment or ‘bottlenecks’, such as replication in different tissues or transmission from individual to individual.

We find that diversity of the quasispecies *per se*, rather than selection of individual adaptive mutations, correlates with enhanced pathogenesis. Our observation of cooperative interactions between different variants within the quasispecies provides a rationale for the role of quasispecies diversity in pathogenesis. For example, certain variants within the population might facilitate the colonization of the gut, another set of mutants might serve as immunological decoys that trick the immune system, and yet another subpopulation might facilitate crossing the blood–brain barrier. Hence, although G64S^{eqs-b} virus re-isolated from brain was highly pathogenic when injected directly into the CNS (Supplementary Fig. S5), it was unable to infect the CNS after intravenous inoculation (Fig. 3a). Maintaining the complexity of the viral quasispecies enables the virus population to spread systemically, perhaps through the complementary functions of different subpopulations, to successfully access the central nervous system.

Taken together, our data support a central concept in quasispecies theory, namely that successful colonization of an ecosystem (in this instance, an infected mouse) occurs by cooperation of different virus variants that occupy distinct regions of the population sequence distribution¹². It is tempting to speculate that this type of positive cooperation also occurs during co-infection of a given host with different viruses, which could have profound consequences for the pathogenic outcome of an infection.

Note added in proof: In Fig. 1c of the advance online publication of this Letter, the LD₅₀ value for G64S^{eqs} appeared incorrectly as 1.7×10^8 . It should be 1.7×10^6 and has been corrected for print.

METHODS

For detailed methods, refer to the Supplementary Information.

Cells and viruses. Tissue culture experiments were performed in HeLa cells (CCL-2.2, ATCC). Wild-type poliovirus type 1 Mahoney was used throughout this study. G64S virus, derived from wild-type virus, contains a Serine in place of Glycine at position 64 of the viral RNA-dependent RNA polymerase. Virus G64S^{sac} contains a silent *SacI* restriction site.

Guanidine hydrochloride resistance (gua^r) assays. The evolution of the viral quasispecies was monitored by the spontaneous emergence of gua^r mutants in each virus population, as previously described¹³.

Artificial expansion of G64S quasispecies by treatment with chemical mutagens. HeLa cells (10^7) were infected with 10^7 plaque-forming units (p.f.u.) of G64S virus in the presence of 400 μ M ribavirin and 125 μ g ml⁻¹ 5-fluorouracil. Mutagenized virus was harvested at total cytopathic effect, and a second round of mutagenesis was performed using the same conditions. After a drop in titre of at least 100-fold (indicating that more than 99% of the genomes had been mutagenized), the population was allowed to recover to normal titre by passaging twice on fresh HeLa cell monolayers in absence of mutagen.

Genomic sequencing for mutational frequency. Using a standard plaque assay, 24 virus isolates from the G64S or wild-type populations were isolated, amplified on HeLa cells, and viral RNA was extracted and purified for RT–PCR. Direct sequencing was performed on PCR products spanning the 5′ non-coding and capsid protein-coding region (nucleotides 480–3300). Additionally, the G64S mutation was confirmed by sequencing to ensure that no reversion to wild-type sequence had taken place. For the G64S^{eqs} population, the entire 3CD precursor sequence, which encodes the poliovirus RNA polymerase, was also sequenced to ensure that no additional mutations had taken place.

Infection of susceptible mice. To determine the 50% lethal dose (LD₅₀), 8-week-old cPVR transgenic mice expressing the poliovirus receptor were inoculated intramuscularly with serial dilutions (20 mice per dilution) of wild-type, G64S or G64S^{eqs} virus. Mice were monitored daily for symptoms leading to total paralysis. LD₅₀ values were determined using the Reed and Muench method. For tissue tropism experiments, mice were inoculated intravenously with 10^7 or 10^8 p.f.u. of wild-type, G64S or G64S^{eqs} virus. Each day after infection, five mice from each group were killed and tissues were collected, homogenized and titred for virus on HeLa cells by standard plaque assay.

To generate virus stocks of brain-passaged virus, brain homogenates from each group (mice infected intravenously with wild-type or G64S^{eqs}, or mice infected intramuscularly with G64S) were pooled, and virus contained therein was amplified once on HeLa cells to reach yields required for re-inoculation experiments. For co-infection experiments, mice were infected intravenously with mixtures of wild-type and G64S^{sac} or G64S^{eqs} and G64S^{sac} viruses. Tissues were collected on day 4. Virus present in tissues was first amplified on HeLa cells, and the amplified viral RNA was used for RT–PCR. The PCR products, spanning the capsid region, were digested with *SacI* and analysed on agarose gel electrophoresis to examine whether wild-type or G64S^{eqs} virus (one band, at approximately 3 kb) or G64S^{sac} virus (two bands, at approximately 1.55 kb and 1.45 kb) were present.

Received 16 September; accepted 25 October 2005.

Published online 4 December 2005.

- Holland, J. *et al.* Rapid evolution of RNA genomes. *Science* **215**, 1577–1585 (1982).
- Holland, J. J., De La Torre, J. C. & Steinhauer, D. A. RNA virus populations as quasispecies. *Curr. Top. Microbiol. Immunol.* **176**, 1–20 (1992).
- Domingo, E. & Holland, J. J. RNA virus mutations and fitness for survival. *Annu. Rev. Microbiol.* **51**, 151–178 (1997).
- Eigen, M. Viral quasispecies. *Sci. Am.* **269**, 42–49 (1993).
- Domingo, E., Holland, J. & Ahlquist, P. *RNA Genetics* (CRC Press, Boca Raton, 1988).
- Domingo, E. & Holland, J. in *Mutations and Rapid Evolution of RNA Viruses* (ed. Morse, S. S.) 161–184 (Raven Press, New York, 1994).
- Domingo, E. Viruses at the edge of adaptation. *Virology* **270**, 251–253 (2000).
- Domingo, E., Sabo, D., Taniguchi, T. & Weissmann, C. Nucleotide sequence heterogeneity of an RNA phage population. *Cell* **13**, 735–744 (1978).
- Coffin, J. M. HIV population dynamics *in vivo*: implications for genetic variation, pathogenesis, and therapy. *Science* **267**, 483–489 (1995).
- Domingo, E. *et al.* Basic concepts in RNA virus evolution. *FASEB J.* **10**, 859–864 (1996).
- Eigen, M. & Biebricher, C. in *RNA Genetics: Variability of RNA Genomes* Vol. 3 (eds Domingo, E., Holland, J. J. & Ahlquist, P.) 211–245 (CRC Press, Boca Raton, 1988).
- Biebricher, C. K. & Eigen, M. The error threshold. *Virus Res.* **107**, 117–127 (2005).
- Crotty, S. *et al.* The broad-spectrum antiviral ribonucleoside ribavirin is an RNA virus mutagen. *Nature Med.* **6**, 1375–1379 (2000).
- Crotty, S., Cameron, C. E. & Andino, R. RNA virus error catastrophe: direct molecular test by using ribavirin. *Proc. Natl Acad. Sci. USA* **98**, 6895–6900 (2001).
- Thompson, A. A. & Peersen, O. B. Structural basis for proteolysis-dependent activation of the poliovirus RNA-dependent RNA polymerase. *EMBO J.* **23**, 3462–3471 (2004).

16. Pfeiffer, J. K. & Kirkegaard, K. A single mutation in poliovirus RNA-dependent RNA polymerase confers resistance to mutagenic nucleotide analogs via increased fidelity. *Proc. Natl Acad. Sci. USA* **100**, 7289–7294 (2003).
17. Baltera, R. F. Jr. & Tershak, D. R. Guanidine-resistant mutants of poliovirus have distinct mutations in peptide 2C. *J. Virol.* **63**, 4441–4444 (1989).
18. Pincus, S. E., Diamond, D. C., Emini, E. A. & Wimmer, E. Guanidine-selected mutants of poliovirus: mapping of point mutations to polypeptide 2C. *J. Virol.* **57**, 638–646 (1986).
19. Arnold, J. J., Vignuzzi, M., Stone, J. K., Andino, R. & Cameron, C. E. Remote-site control of an active-site fidelity checkpoint in a viral RNA-dependent RNA polymerase. *J. Biol. Chem.* **280**, 25706–25716 (2005).
20. Gohara, D. W., Arnold, J. J. & Cameron, C. E. Poliovirus RNA-dependent RNA polymerase (3D^{pol}): kinetic, thermodynamic, and structural analysis of ribonucleotide selection. *Biochemistry* **43**, 5149–5158 (2004).
21. Ren, R. & Racaniello, V. R. Poliovirus spreads from muscle to the central nervous system by neural pathways. *J. Infect. Dis.* **166**, 747–752 (1992).
22. Anderson, J. P., Daifuku, R. & Loeb, L. A. Viral error catastrophe by mutagenic nucleosides. *Annu. Rev. Microbiol.* **58**, 183–205 (2004).
23. Pariente, N., Sierra, S., Lowenstein, P. R. & Domingo, E. Efficient virus extinction by combinations of a mutagen and antiviral inhibitors. *J. Virol.* **75**, 9723–9730 (2001).
24. Pariente, N., Airaksinen, A. & Domingo, E. Mutagenesis versus inhibition in the efficiency of extinction of foot-and-mouth disease virus. *J. Virol.* **77**, 7131–7138 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to J. Frydman, D. Ganem, A. Frankel and members of the Andino laboratory for critical reading of the manuscript. This work was supported by NIH-NIAID grants to R.A. and C.E.C. and a predoctoral NIH fellowship to J.K.S.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to R.A. (andino@itsa.ucsf.edu).

Oscillations of cyclic AMP in hormone-stimulated insulin-secreting β -cells

Oleg Dyachok^{1,2}, Yegor Isakov¹, Jenny S  getorp¹ & Anders Tengholm¹

Cyclic AMP is a ubiquitous second messenger that transduces signals from a variety of cell surface receptors to regulate diverse cellular functions, including secretion, metabolism and gene transcription. In pancreatic β -cells, cAMP potentiates Ca^{2+} -dependent exocytosis^{1–3} and mediates the stimulation of insulin release exerted by the hormones glucagon and glucagon-like peptide-1 (GLP-1) (refs 4–6). Whereas Ca^{2+} signals have been extensively characterized and shown to involve oscillations important for the temporal control of insulin secretion^{4,7,8}, the kinetics of receptor-triggered cAMP signals is unknown. Here we introduce a new ratiometric evanescent-wave-microscopy approach to measure cAMP concentration beneath the plasma membrane, and show that insulin-secreting β -cells respond to glucagon and GLP-1 with marked cAMP oscillations. Simultaneous measurements of intracellular Ca^{2+} concentration revealed that the two messengers are interlinked and reinforce each other. Moreover, cAMP oscillations are capable of inducing rapid on–off Ca^{2+} responses, but only sustained elevation of cAMP concentration induces nuclear translocation of the catalytic subunit of the cAMP-dependent protein kinase. Our results establish a new signalling mode for cAMP and indicate that temporal encoding of cAMP signals might constitute a basis for differential regulation of downstream cellular targets.

To investigate the kinetics of hormone-evoked cAMP signals in β -cells we developed a fluorescent biosensor that reports cAMP concentration beneath the plasma membrane ($[\text{cAMP}]_i$). A truncated form of the regulatory subunit of the cAMP-dependent protein kinase (PKA) was labelled with cyan fluorescent protein (CFP) and targeted to the plasma membrane with a polybasic sequence and a farnesylation motif ($\Delta\text{RII}\beta$ -CFP-CAAX; Supplementary Fig. S1). Because of its high affinity for $\Delta\text{RII}\beta$ -CFP-CAAX, co-expressed PKA catalytic subunit (C α) labelled with yellow fluorescent protein (YFP) also located to the plasma membrane. Holoenzyme dissociation as $[\text{cAMP}]_i$ increases could then be monitored as C α -YFP redistribution to the cytoplasm (Supplementary Fig. S1). With the use of evanescent wave microscopy⁹, such translocation results in large fluorescence changes that can be measured with minimal photobleaching and phototoxic effects and a signal-to-noise ratio superior to that of other fluorescence microscopy approaches¹⁰.

Inhibition of phosphodiesterases with 100 μM 3-isobutyl-1-methylxanthine (IBMX) or activation of adenylyl cyclases with 5 μM forskolin in INS-1 β -cells expressing the biosensor did not affect evanescent-wave-excited $\Delta\text{RII}\beta$ -CFP-CAAX fluorescence, but induced a prompt loss of membrane C α -YFP fluorescence that was rapidly reversed after washout of the drugs (Fig. 1a). To compensate for fluorescence changes that might occur independently of $[\text{cAMP}]_i$, such as changes in cell morphology or adhesion, we calculated the ratio between the $\Delta\text{RII}\beta$ -CFP-CAAX and C α -YFP signals ($R_{\text{CFP/YFP}}$) and normalized the prestimulatory level to 1 (Fig. 1a). The response

kinetics, measured as time to half-maximal change ($t_{1/2}$) of $R_{\text{CFP/YFP}}$, averaged 23 ± 2 s (mean $\pm$ s.e.m; $n = 23$) and 20 ± 3 s ($n = 15$) after stimulation with IBMX and forskolin, respectively. After removal of the drugs, $R_{\text{CFP/YFP}}$ returned to the prestimulatory level with $t_{1/2} = 14 \pm 3$ s ($n = 14$; IBMX) and 27 ± 3 s ($n = 7$; forskolin). Direct bath application of 10 μM cAMP in cells permeabilized with staphylococcal α -toxin resulted in even faster C α -YFP translocation, with $t_{1/2} = 11.7 \pm 0.9$ s ($n = 11$; Supplementary Fig. S2). The specificity of probe dissociation was verified by exposing intact cells to membrane-permeable cyclic nucleotide analogues. Whereas 10 mM 8-bromo-cGMP lacked effect, the same concentration of 8-bromo-cAMP induced a stable $82 \pm 9\%$ ($n = 5$) increase in $R_{\text{CFP/YFP}}$ (Supplementary Fig. S2).

Next, we investigated $[\text{cAMP}]_i$ changes induced by physiological stimuli. In pancreatic β -cells, the gut-derived peptide hormone GLP-1 is an important stimulator of insulin secretion, cell development, growth and survival^{4,11}. It exerts its action through binding to G-protein-coupled receptors and the activation of adenylyl cyclases. Stimulation of rat insulinoma cells by using 10 nM GLP-1 resulted in a rapid and pronounced increase in $[\text{cAMP}]_i$ ($t_{1/2} = 24 \pm 3$ s; $n = 20$) that was reversed when the hormone was removed (Fig. 1b). Increasing the glucose concentration from 3 mM to 20 mM had a negligible effect on $[\text{cAMP}]_i$ by its own, but markedly enhanced the response to GLP-1 (Fig. 1b; the response at 20 mM glucose was $198 \pm 29\%$ of that at 3 mM glucose; $n = 7$; $P < 0.02$), which is consistent with the previously described synergism between glucose and GLP-1 on cAMP production¹². The GLP-1-induced $[\text{cAMP}]_i$ elevation was rapidly reversed on inhibition of adenylyl cyclases by 10 μM of the α_2 -adrenergic receptor agonist noradrenaline ($88 \pm 2\%$ suppression; $t_{1/2} = 12 \pm 2$ s; $n = 6$; $P < 0.001$; Fig. 1c). Noradrenaline also reversed the elevation of $[\text{cAMP}]_i$ induced by 100 μM IBMX (Supplementary Fig. S3). The average $t_{1/2}$ of 59 ± 3 s ($n = 9$, $P < 0.001$) was consistent with retarded cAMP hydrolysis in the presence of the phosphodiesterase inhibitor. Application of noradrenaline under resting conditions resulted in a modest decrease in $R_{\text{CFP/YFP}}$ showing that the biosensor detects even small changes in basal $[\text{cAMP}]_i$ and that $R_{\text{CFP/YFP}}$ in non-stimulated cells was as high as $7.4 \pm 1.3\%$ ($n = 6$) or $7.5 \pm 0.9\%$ ($n = 6$) of that reached after treatment with 10 nM GLP-1 or 100 μM IBMX, respectively (Supplementary Fig. S3). From the rapid and pronounced changes in $[\text{cAMP}]_i$ induced by IBMX, GLP-1 and noradrenaline we conclude that there are high rates of cAMP production and degradation beneath the plasma membrane.

High turnover of cAMP is favourable for establishing spatial gradients or oscillations of $[\text{cAMP}]_i$. Indeed, INS-1 β -cells stimulated with 0.3–1.0 nM GLP-1 frequently responded with $[\text{cAMP}]_i$ oscillations from the baseline or a slightly elevated level (43 of 51 cells; Fig. 1d–h; Supplementary Fig. S4). An increase in the GLP-1 concentration typically resulted in dose-dependent sustained

¹Department of Medical Cell Biology, Uppsala University, BMC, Box 571, SE-751 23 Uppsala, Sweden. ²Department of Biophysics, Kiev T. Shevchenko National University, 01033 Kiev, Ukraine.

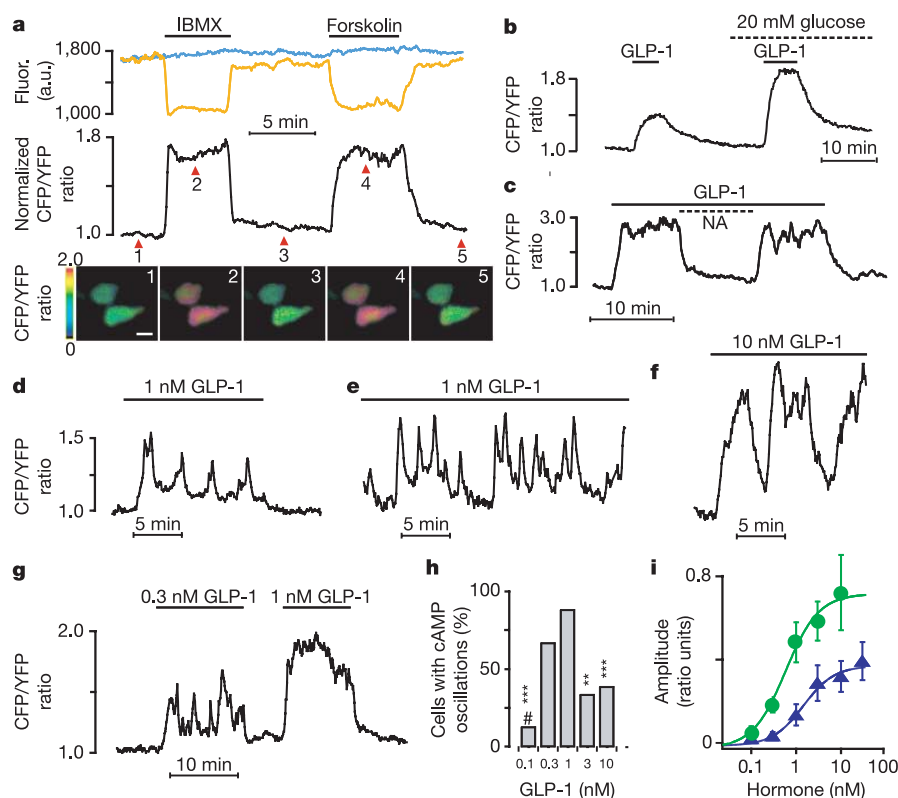


Figure 1 | Ratiometric evanescent-wave-microscopy recordings of [cAMP]_i in individual insulin-secreting β -cells. **a**, Δ R11 β -CFP-CAAX (blue) and C α -YFP (yellow) fluorescence recorded from an INS-1 β -cell during stimulation with 100 μ M IBMX and with 5 μ M forskolin. The CFP/YFP fluorescence ratio (black) was normalized to the prestimulatory level. Pseudocoloured ratio images are from the time points indicated with numbered arrowheads. Scale bar, 10 μ m. **b**, Effect of 10 nM GLP-1 and elevation of glucose from 3 to 20 mM in an individual INS-1E β -cell. **c**, [cAMP]_i responses to 1 nM GLP-1 and 10 μ M noradrenaline (NA).

d–f, Examples of GLP-1-induced [cAMP]_i oscillations showing different amplitudes and frequencies. **g**, Concentration-dependent switch of [cAMP]_i response pattern. **h**, Percentage oscillating cells at different GLP-1 concentrations. Values of n : 8 (0.1 nM), 9 (0.3 nM), 42 (1 nM), 9 (3 nM) and 26 (10 nM). Three asterisks, $P < 0.001$; two asterisks, $P < 0.01$ for differences in proportions from 1 nM GLP-1; hash sign, $P < 0.05$ for differences from 0.3 nM GLP-1. **i**, Dose dependence of [cAMP]_i elevations induced by glucagon (blue) and GLP-1 (green). Results are means $\pm$ s.e.m. for two to eight cells at each concentration.

[cAMP]_i elevation (Fig. 1g), and at hormone concentrations of 3–10 nM, oscillations occurred in less than 40% of the cells ($n = 35$; Fig. 1f, h). The oscillatory patterns varied between different cells with frequencies ranging from 0.16 to 1.5 min⁻¹ and amplitudes from 0.08 to 1.7 normalized $R_{\text{CFP/YFP}}$ units. Similar experiments with 0.3–30 nM glucagon demonstrated that this hormone also could elicit oscillatory [cAMP]_i responses, although in a much smaller proportion of the cells (11 of 50 cells; Supplementary Fig. S4). Glucagon was also less efficient than GLP-1 in elevating [cAMP]_i. The maximal average amplitudes of [cAMP]_i elevation were 0.39 ± 0.09 ratio units ($n = 6$) for glucagon and 0.72 ± 0.18 units ($n = 4$) for GLP-1, with half-maximal responses achieved at 1.51 ± 0.34 nM glucagon and 0.65 ± 0.15 nM GLP-1 ($P < 0.05$ for difference in half-maximal concentration; Fig. 1i). These EC₅₀ values are almost identical to those previously obtained with conventional radiotracer techniques^{13,14}.

cAMP-generating stimuli are known to induce intracellular Ca²⁺ concentration ([Ca²⁺]_i) signals in insulin-secreting β -cells by the PKA-mediated modulation of K_{ATP} channels, voltage-gated Ca²⁺ channels and Ins(1,4,5)P₃ receptors^{4,5,15}. Accordingly, INS-1 β -cells responded to 1–10 nM GLP-1 with pronounced oscillations of [Ca²⁺]_i (average amplitude 562 ± 52 nM, frequency range 0.19–2.52 min⁻¹, $n = 31$ cells; Fig. 2a). It is also well known that the formation and breakdown of cAMP can be affected by [Ca²⁺]_i through direct or indirect effects on adenylyl cyclases and phosphodiesterases^{12,16–18}. The [cAMP]_i oscillations induced by 1 nM GLP-1 immediately disappeared when extracellular Ca²⁺ was removed and the Ca²⁺ chelator EGTA was added at 2 mM (Fig. 2b; $n = 3$). This

effect was reversible with restoration of the oscillations when Ca²⁺ was reintroduced. Interactions between Ca²⁺ and cAMP at multiple levels have been proposed to cause [cAMP]_i oscillations in some systems^{19,20}. Depending on the particular isoforms of adenylyl cyclases and phosphodiesterases expressed, elevation of [Ca²⁺]_i can either increase or decrease [cAMP]_i. To investigate whether the [cAMP]_i oscillations in β -cells are synchronized with [Ca²⁺]_i signals we performed simultaneous measurements of [Ca²⁺]_i and [cAMP]_i. As shown in Fig. 2c, elevations of [cAMP]_i coincided with those of [Ca²⁺]_i ($n = 3$) and this coordination was confirmed by cross-correlation analysis (Supplementary Fig. S5). Although the initial [cAMP]_i elevation preceded that of [Ca²⁺]_i on the addition of GLP-1, there was no detectable time difference during the subsequent oscillations (Fig. 2d). Thus, [cAMP]_i and [Ca²⁺]_i signals are temporally coordinated in insulin-secreting cells. This synchronization and the mutual reinforcement of cAMP and Ca²⁺ signals should constitute an exquisite trigger of exocytosis and might explain how GLP-1 selectively enhances the pulsatile component of insulin secretion in healthy and diabetic subjects^{21,22}.

Oscillatory cAMP signalling involving the excretion of cAMP and the activation of an extracellular cAMP receptor is previously known from the slime mould *Dictyostelium*²³, and spontaneous cAMP transients in *Xenopus* frog embryonal neurons were recently reported²⁴. Our present data provide the first demonstration that activation of hormone receptors can evoke [cAMP]_i oscillations. This is significant because signalling with oscillations might help to improve low-level signal detection and achieve specificity in downstream effects, as has been described for Ca²⁺ (refs 19, 25). To test

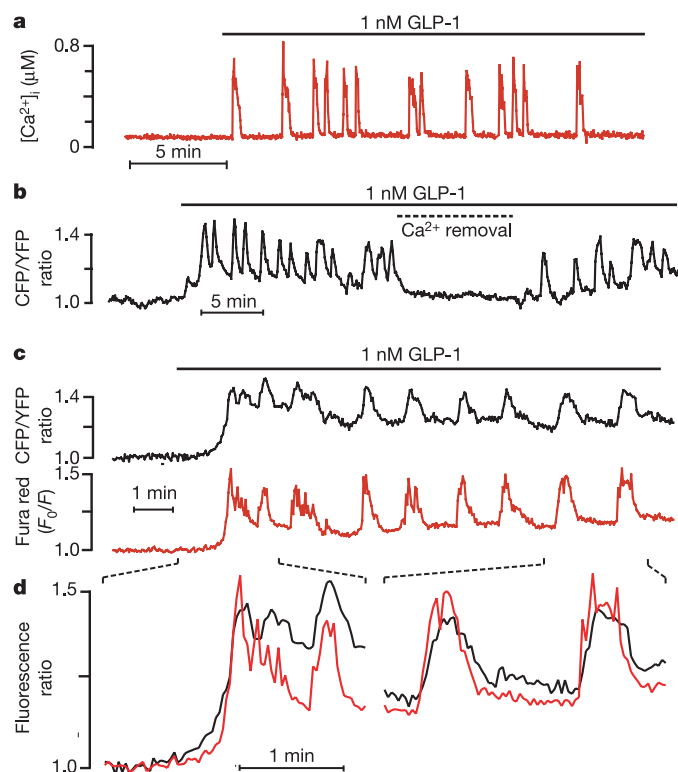


Figure 2 | Interdependence of $[cAMP]_i$ and $[Ca^{2+}]_i$ oscillations in individual INS-1 β -cells. **a**, $[Ca^{2+}]_i$ oscillations in a Fura-2-loaded cell stimulated with 1 nM GLP-1. **b**, Suppression of GLP-1-induced $[cAMP]_i$ oscillations by the removal of extracellular Ca^{2+} and the addition of 2 mM EGTA. **c**, Simultaneous recording of $[cAMP]_i$ (black) and $[Ca^{2+}]_i$ (red) in a Fura-red-loaded cell stimulated with 1 nM GLP-1. **d**, Time-expanded portions of the traces in **c**, showing $[cAMP]_i$ (black) and $[Ca^{2+}]_i$ (red) plotted with common axes.

whether different temporal patterns of cAMP signals might contribute to the selective regulation of downstream events we evaluated the effects of stable and oscillatory cAMP signals on two well-established cAMP-dependent responses, the generation of $[Ca^{2+}]_i$ oscillations (see above) and the translocation of the PKA catalytic subunit to the nucleus²⁶. As expected, continuous stimulation with 100 μ M IBMX resulted in stable elevation of $[cAMP]_i$ (Fig. 3a), whereas application of IBMX in pulses generated regular $[cAMP]_i$ oscillations (Fig. 3b). Both stimulation protocols promoted the appearance of $[Ca^{2+}]_i$ oscillations, and with pulsatile stimulation the $[Ca^{2+}]_i$ signals coincided precisely with the presence of IBMX ($n = 37$; Fig. 3c, d).

Because the YFP-tagged PKA catalytic subunit is too large to enter the nucleus, we assessed cAMP-induced PKA nuclear translocation by tagging the $C\alpha$ subunit with a small tetracysteine motif ($C\alpha$ -Cys₄; Supplementary Fig. S6) that can be specifically labelled with the membrane-permeable fluorescent biarsenical dye FAsH (ref. 27). Epifluorescence imaging of FAsH-labelled $C\alpha$ -Cys₄ revealed cytoplasmic distribution with markedly less nuclear fluorescence in non-stimulated cells and translocation of the construct into the nucleus after elevation of $[cAMP]_i$ (Supplementary Fig. S6). The ratio of nuclear to cytoplasmic fluorescence increased from 0.66 ± 0.02 ($n = 77$) in non-stimulated cells to 0.90 ± 0.03 ($n = 65$; $P < 0.001$) after 25 min of stable $[cAMP]_i$ elevation induced by 100 μ M IBMX (Fig. 3e, f). In contrast, $[cAMP]_i$ oscillations generated by pulsatile IBMX (1 min stimulation and 3 min wash) failed to induce nuclear translocation of PKA (0.68 ± 0.02 , $n = 63$) even after more than 100 min when the cells had been exposed to an equivalent total dose of IBMX. The strikingly higher efficiency of stable compared with oscillatory $[cAMP]_i$ elevation to induce the nuclear translocation of PKA was confirmed in experiments with continuous

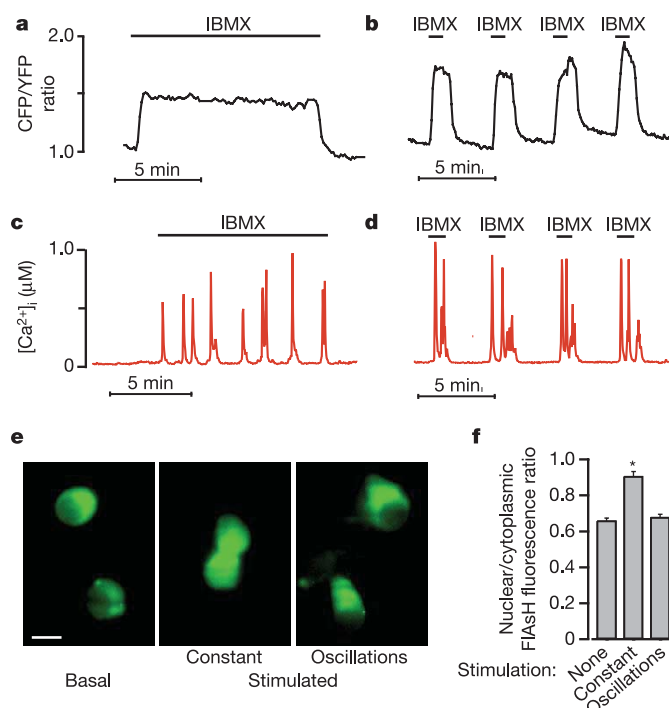


Figure 3 | Effect of $[cAMP]_i$ kinetics on $[Ca^{2+}]_i$ and nuclear translocation of PKA catalytic subunit. **a**, **b**, $[cAMP]_i$ signals in individual INS-1 β -cells induced by continuous (**a**) and pulsatile (**b**) application of 100 μ M IBMX. **c**, **d**, $[Ca^{2+}]_i$ oscillations evoked by stable (**c**) and oscillatory (**d**) $[cAMP]_i$ elevation in individual Fura-2-loaded cells. **e**, Distribution of FAsH-labelled $C\alpha$ -Cys₄ in INS-1 β -cells under basal conditions and after 25 min of continuous or oscillatory elevation of $[cAMP]_i$. **f**, Nuclear/cytoplasmic FAsH fluorescence ratios (means $\pm$ s.e.m.) in cells exposed to different patterns of $[cAMP]_i$ elevation. $n = 77$, 65 and 63 cells for basal, continuous and oscillatory stimulation, respectively. Asterisk, $P < 0.001$ for difference from basal.

stimulation with 25 μ M IBMX (0.89 ± 0.05 , $n = 20$, $P < 0.001$), providing a time-average dose identical to 100 μ M pulsatile IBMX.

Thus, hormone receptor activation in β -cells can evoke cAMP oscillations beneath the plasma membrane. On the basis of the present findings, we suggest that brief transients of $[cAMP]_i$ are important in the selective regulation of rapid local cytoplasmic events, such as ion channel activity and exocytosis, whereas prolonged $[cAMP]_i$ elevation is required for long-term effects such as PKA-mediated activation of the nuclear transcription factors involved in regulating cell survival and proliferation.

METHODS

cDNA constructs. Δ RII β -CFP-CAAX was created by ligating a cDNA fragment encoding residues 81–416 of rat PKA RII β to the amino terminus of CFP, which was targeted to the plasma membrane with a polybasic sequence and a CAAX motif from human Ki-Ras. $C\alpha$ -YFP was generated by fusing the full-length coding sequence of mouse PKA $C\alpha$ to the N terminus of YFP. $C\alpha$ -Cys₄ was made by ligating $C\alpha$ cDNA into the pENTR4 Gateway entry vector (Invitrogen) followed by site-specific recombination into the pcDNA6.2-Lumio vector (Invitrogen).

Cell culture and transfection. Insulin-secreting INS-1 cells (passage 90–120; ref. 28) and the more glucose-responsive subclone INS-1E (ref. 29; used for the experiment in Fig. 1b), were cultured in RPMI 1640 medium (Invitrogen) containing 11 mM glucose and supplemented with 10% fetal calf serum (Invitrogen), 1 mM sodium pyruvate, 2 mM glutamine, 50 μ M mercaptoethanol, 100 U ml⁻¹ penicillin and 100 μ g ml⁻¹ streptomycin. After plating on coverslips, the cells were transfected with Lipofectamine 2000 (Invitrogen) in accordance with the manufacturer's protocol.

Imaging. At 12–24 h after transfection the cells were transferred to buffer containing 138 mM NaCl, 4.8 mM KCl, 1.3 mM CaCl₂, 1.2 mM MgCl₂, 3 mM glucose and 25 mM HEPES, pH 7.40. Where indicated, the cells were loaded with the Ca^{2+} indicators Fura-2 or Fura red by incubation for 30 or 40 min at 37 °C

with their acetoxymethyl esters at 1 or 5 μM , respectively. $\text{C}\alpha\text{-Cys}_4$ -transfected cells were labelled by incubation at 37 °C for 20 min with 0.5 μM FAsH-EDT₂ (Lumio Green labelling reagent; Invitrogen). These cells were then rinsed with buffer, incubated for 10–15 min in 250 μM EDT and again rinsed with buffer to reduce non-specific labelling. It was ascertained that background FAsH labelling was unaffected by continuous $[\text{cAMP}]_i$ elevation in non-transfected cells, although it probably contributed to an underestimation of the cAMP-induced change in the FAsH nuclear/cytoplasmic fluorescence ratio in cells expressing $\text{C}\alpha\text{-Cys}_4$.

$[\text{Ca}^{2+}]_i$ imaging of Fura-2-loaded cells was performed as described previously¹⁵. Translocation of the cAMP reporter was measured using a custom-built evanescent wave microscope equipped with a 60 $\times$ objective with a numerical aperture of 1.45 (Nikon) and an Orca-ER camera (Hamamatsu) controlled by MetaFluor software (Molecular Devices). The same setup was used for epifluorescence imaging. Excitation and emission wavelengths were selected with the following filters: CFP, 458 nm/10 nm half-bandwidth and 485/25 nm; YFP 514.5/10 and 550/30 nm; Fura red 514.5/10 and 630 nm long-pass; FAsH 488/10 and 530/50 nm. Exposure times were from 100 to 400 ms and images were acquired every 2–5 s.

Statistics. Results are reported as means $\pm$ s.e.m. Statistical differences were evaluated with Student's *t*-test or Fisher's exact test (Fig. 1h).

Received 26 July; accepted 8 November 2005.

- Prentki, M. & Matschinsky, F. M. Ca^{2+} , cAMP, and phospholipid-derived messengers in coupling mechanisms of insulin secretion. *Physiol. Rev.* **67**, 1185–1248 (1987).
- Renström, E., Eliasson, L. & Rorsman, P. Protein kinase A-dependent and -independent stimulation of exocytosis by cAMP in mouse pancreatic B-cells. *J. Physiol. (Lond.)* **502**, 105–118 (1997).
- Ozaki, N. *et al.* cAMP-GEFII is a direct target of cAMP in regulated exocytosis. *Nature Cell Biol.* **2**, 805–811 (2000).
- Gromada, J., Brock, B., Schmitz, O. & Rorsman, P. Glucagon-like peptide-1: regulation of insulin secretion and therapeutic potential. *Basic Clin. Pharmacol. Toxicol.* **95**, 252–262 (2004).
- Holz, G., Kühtreiber, W. & Habener, J. Pancreatic beta-cells are rendered glucose-competent by the insulinotropic hormone glucagon-like peptide-1(7–37). *Nature* **361**, 362–365 (1993).
- Eliasson, L. *et al.* SUR1 regulates PKA-independent cAMP-induced granule priming in mouse pancreatic B-cells. *J. Gen. Physiol.* **121**, 181–197 (2003).
- Gilon, P., Shepherd, R. & Henquin, J. C. Oscillations of secretion driven by oscillations of cytoplasmic Ca^{2+} as evidences in single pancreatic islets. *J. Biol. Chem.* **268**, 22265–22268 (1993).
- Bergsten, P., Grapenhiesser, E., Gylfe, E., Tengholm, A. & Hellman, B. Synchronous oscillations of cytoplasmic Ca^{2+} and insulin release in glucose-stimulated pancreatic islets. *J. Biol. Chem.* **269**, 8749–8753 (1994).
- Steyer, J. A. & Almers, W. A real-time view of life within 100 nm of the plasma membrane. *Nature Rev. Mol. Cell Biol.* **2**, 268–275 (2001).
- Tengholm, A., Teruel, M. N. & Meyer, T. Single cell imaging of PI3K activity and glucose transporter insertion into the plasma membrane by dual colour evanescent wave microscopy. *Sci. STKE* PL4 (2003).
- List, J. F. & Habener, J. Glucagon-like peptide 1 agonists and the development and growth of pancreatic β -cells. *Am. J. Physiol. Endocrinol. Metab.* **286**, E875–E881 (2004).
- Delmeire, D. *et al.* Type VIII adenylyl cyclase in rat beta cells: coincidence signal detector/generator for glucose and GLP-1. *Diabetologia* **46**, 1383–1393 (2003).
- Widmann, C., Bürki, E., Dolci, W. & Thorens, B. Signal transduction by the cloned glucagon-like peptide-1 receptor: comparison with signalling by the endogenous receptors of β cell lines. *Mol. Pharmacol.* **45**, 1029–1035 (1994).
- Ma, X. *et al.* Glucagon stimulates exocytosis in mouse and rat pancreatic α -cells by binding to glucagon receptors. *Mol. Endocrinol.* **19**, 198–212 (2005).
- Dyachok, O. & Gylfe, E. Ca^{2+} -induced Ca^{2+} release via inositol 1,4,5-trisphosphate receptors is amplified by protein kinase A and triggers exocytosis in pancreatic β -cells. *J. Biol. Chem.* **279**, 45455–45461 (2004).
- Pyne, N. J. & Furman, B. L. Cyclic nucleotide phosphodiesterases in pancreatic islets. *Diabetologia* **46**, 1179–1189 (2003).
- Cooper, D. Regulation and organization of adenylyl cyclases and cAMP. *Biochem. J.* **375**, 517–529 (2003).
- Mehats, C., Andersen, C. B., Filopanti, M., Jin, S.-L. C. & Conti, M. Cyclic nucleotide phosphodiesterases and their role in endocrine cell signalling. *Trends Endocrinol. Metab.* **13**, 29–35 (2002).
- Cooper, D. M. F., Mons, N. & Karpen, J. W. Adenylyl cyclases and the interaction between calcium and cAMP signalling. *Nature* **374**, 421–424 (1995).
- Rich, T. C. & Karpen, J. W. Review article: cyclic AMP sensors in living cells: what signals can they actually measure? *Ann. Biomed. Eng.* **30**, 1088–1099 (2002).
- Pørksen, N. *et al.* Glucagon-like peptide increases mass but not frequency or orderliness of pulsatile insulin secretion. *Diabetes* **47**, 45–49 (1998).
- Ritzel, R. *et al.* Glucagon-like peptide 1 increases secretory burst mass of pulsatile insulin secretion in patients with type 2 diabetes and impaired glucose tolerance. *Diabetes* **50**, 776–784 (2001).
- Maeda, M. *et al.* Periodic signalling controlled by an oscillatory circuit that includes protein kinases ERK2 and PKA. *Science* **304**, 875–878 (2004).
- Gorbunova, Y. V. & Spitzer, N. C. Dynamic interactions of cyclic AMP transients and spontaneous Ca^{2+} spikes. *Nature* **418**, 93–96 (2002).
- Dolmetsch, R. E., Xu, K. & Lewis, R. S. Calcium oscillations increase the efficiency and specificity of gene expression. *Nature* **392**, 933–936 (1998).
- Harootunian, A. T. *et al.* Movement of the free catalytic subunit of cAMP-dependent protein kinase into and out of the nucleus can be explained by diffusion. *Mol. Biol. Cell* **4**, 993–1002 (1993).
- Griffin, B. A., Adams, S. R. & Tsien, R. Y. Specific covalent labeling of recombinant protein molecules inside living cells. *Science* **281**, 269–272 (1998).
- Asfari, M. *et al.* Establishment of 2-mercaptoethanol-dependent differentiated insulin-secreting cell lines. *Endocrinology* **130**, 167–178 (1992).
- Merglen, A. *et al.* Glucose-sensitivity and metabolism-secretion coupling studied during two-year continuous culture of INS-1E insulinoma cells. *Endocrinology* **145**, 667–678 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank S. McKnight for providing cDNA for the regulatory and catalytic subunits of PKA; C. Wollheim and P. Maeckler for the rat insulinoma cells; P. Korzhavii for assistance with cross-correlation analysis; and E. Gylfe, C.-H. Heldin, B. Hellman and T. Meyer for reading the manuscript. This study was supported by grants from Åke Wiberg's Foundation, Carl Trygger's Foundation for Scientific Research, the European Foundation for the Study of Diabetes/Novo Nordisk, the Family Ernfor Foundation, Novo Nordisk Foundation, Swedish Diabetes Association, Swedish Research Council, and the Wenner-Gren Foundations.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.T. (anders.tengholm@medcellbiol.uu.se).

Oncogenic pathway signatures in human cancers as a guide to targeted therapies

Andrea H. Bild^{1,2}, Guang Yao^{1,2}, Jeffrey T. Chang^{1,2}, Quanli Wang¹, Anil Potti^{1,4}, Dawn Chasse^{1,2}, Mary-Beth Joshi³, David Harpole³, Johnathan M. Lancaster⁷, Andrew Berchuck⁵, John A. Olson Jr^{1,3}, Jeffrey R. Marks³, Holly K. Dressman^{1,2}, Mike West⁶ & Joseph R. Nevins^{1,2}

The development of an oncogenic state is a complex process involving the accumulation of multiple independent mutations that lead to deregulation of cell signalling pathways central to the control of cell growth and cell fate^{1–3}. The ability to define cancer subtypes, recurrence of disease and response to specific therapies using DNA microarray-based gene expression signatures has been demonstrated in multiple studies⁴. Various studies have also demonstrated the potential for using gene expression profiles for the analysis of oncogenic pathways^{5–11}. Here we show that gene expression signatures can be identified that reflect the activation status of several oncogenic pathways. When evaluated in several large collections of human cancers, these gene expression signatures identify patterns of pathway deregulation in tumours and clinically relevant associations with disease outcomes. Combining signature-based predictions across several pathways identifies coordinated patterns of pathway deregulation that distinguish between specific cancers and tumour subtypes. Clustering tumours based on pathway signatures further defines prognosis in respective patient subsets, demonstrating that patterns of oncogenic pathway deregulation underlie the development of the oncogenic phenotype and reflect the biology and outcome of specific cancers. Predictions of pathway deregulation in cancer cell lines are also shown to predict the sensitivity to therapeutic agents that target components of the pathway. Linking pathway deregulation with sensitivity to therapeutics that target components of the pathway provides an opportunity to make use of these oncogenic pathway signatures to guide the use of targeted therapeutics.

We used human primary mammary epithelial cell cultures (HMECs) to develop a series of pathway signatures. Recombinant adenoviruses were used to express various oncogenic activities in an otherwise quiescent cell, thereby specifically isolating the subsequent events as defined by the activation/deregulation of that single pathway. Various biochemical measures demonstrate pathway activation (Supplementary Fig. 1). RNA from multiple independent infections was collected for DNA microarray analysis using Affymetrix Human Genome U133 Plus 2.0 Array. Gene expression signatures that reflect the activity of a given pathway are identified using supervised classification methods of analysis previously described¹². The analysis selects a set of genes for which the expression levels are most highly correlated with the classification of HMEC samples into oncogene-activated/deregulated versus control (green fluorescent protein, GFP). The dominant principal components from such a set of genes then defines a relevant phenotype-related metagene, and regression models assign the relative probability of pathway deregulation in tumour or cell line samples.

It is clear from Fig. 1a that the various signatures distinguish cells expressing the oncogenic activity from control cells. Given the potential for overlap in the pathways, we also examined the extent to which the signatures distinguish one pathway from another. Use of the first three principal components from each signature, evaluated across all experimental samples, demonstrates that the patterns of expression in each signature are specific to each pathway; the gene expression patterns accurately distinguish the individual oncogenic effects despite overlapping downstream consequences (Fig. 1b). The genes identified as comprising each signature are listed in Supplementary Table 1. To evaluate more formally the predictive validity and robustness of the pathway signatures, a leave-one-out cross validation study was applied to the set of pathway predictors. This analysis demonstrates that these signatures of oncogenic pathways can accurately predict the cells expressing the oncogenic activity from the control cells (Supplementary Fig. 2). The analysis clearly distinguishes and predicts the state of an oncogenic pathway.

Further verification of the capacity of oncogenic pathway signatures to predict accurately the status of pathways made use of tumour samples derived from various mouse cancer models. Pathway signatures were regenerated from the genes common to both human and mouse data sets; the analysis was trained on the HMEC-derived signatures and then used to predict the pathway status of all tumours. These studies were carried out using three of the pathway signatures for which we had matching mouse models that could be used for validation: Myc, Ras and E2F3. Across the set of mouse tumours, this analysis evaluates the relative probability of pathway deregulation of each tumour—that is, the predicted status of the pathway in each mouse tumour based only on the signatures developed in HMECs. These predictions are displayed as a colour map: red indicates a high probability of pathway deregulation and blue indicates a low probability, with predictions sorted by the relative probability of pathway deregulation. As shown in Fig. 2a, the pathway predictions exhibit close correlation with the molecular basis for tumour induction. For instance, the five mouse mammary tumour virus (MMTV)-MYC tumours exhibit the highest probability of Myc pathway deregulation, whereas the six Rb null tumours exhibit the highest probability of E2F3 deregulation. The probability of Ras pathway activation was highest in the MMTV-HRAS animals and MMTV-MYC tumours; this indication of Ras pathway activation in the MMTV-MYC tumours is consistent with past results demonstrating a selection for Ras mutations in these tumours^{6,13}. Further substantiation and validation was obtained from a series of tumours in which Ras activity was spontaneously activated by homologous recombination in adult animals, more closely

¹Institute for Genome Sciences and Policy, Duke University, Durham, North Carolina 27708, USA. ²Department of Molecular Genetics and Microbiology, ³Department of Surgery, ⁴Department of Medicine, ⁵Department of Obstetrics & Gynecology, Duke University Medical Center, Durham, North Carolina 27710, USA. ⁶Institute of Statistics and Decision Sciences, Duke University, Durham, North Carolina 27708, USA. ⁷H. Lee Moffitt Cancer Center & Research Institute, University of South Florida, Tampa, Florida 33612, USA.

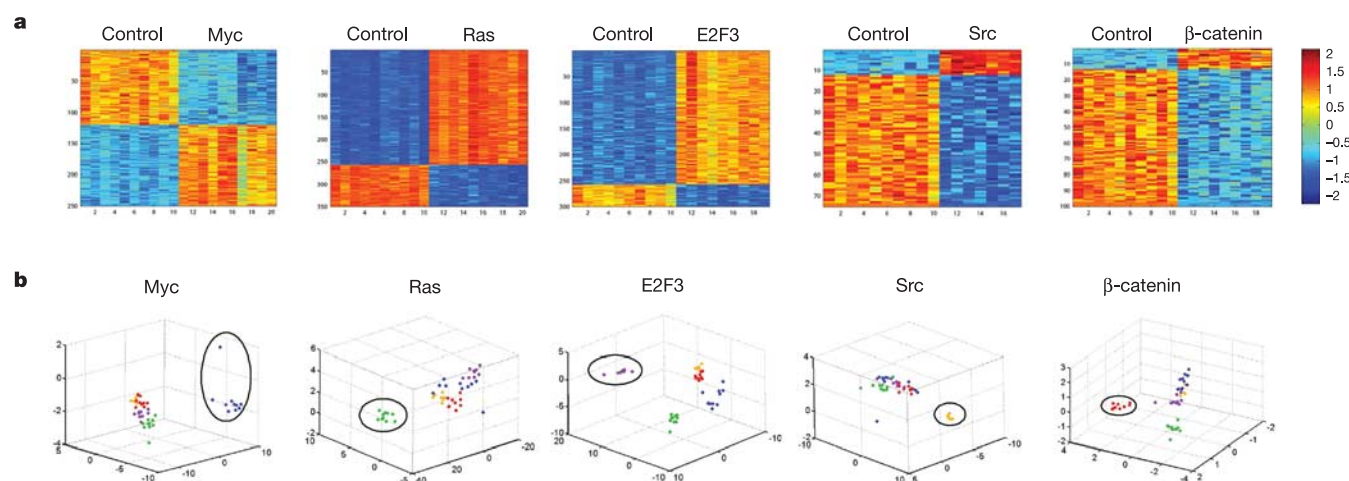


Figure 1 | Gene expression patterns that predict oncogenic pathway deregulation. **a**, Image intensity display of the expression levels of genes most highly weighted in the predictor differentiating GFP-expressing control cells from cells expressing the indicated oncogenic activity. Expression levels are standardized to zero mean and unit variance across samples, displayed with genes as rows and samples as columns, and colour coded to indicate high (red) or low (blue) expression levels. **b**, Scatter plots

depicting the classification of samples based on the first three principal components (expression patterns) derived from each signature, as shown in **a**. The gene expression values for each signature were extracted from all experimental samples and mean centred, then single value decomposition analysis was applied across all samples. Colour coding for samples is as follows: Myc, blue; Ras, green; E2F3, purple; Src, yellow; β-catenin, red. Samples representing the specific pathway being examined are circled.

mimicking pathway deregulation in human tumours¹⁴. There was a consistent prediction of Ras pathway deregulation within these tumours when compared to the set of samples from control lung tissue (Fig. 2b). Taken together, these results strongly support the conclusion that the various oncogenic pathway signatures do reliably reflect pathway status under a variety of circumstances, and thus can serve as useful tools to probe the status of these pathways.

Previous work has linked Ras activation with the development of adenocarcinomas of the lung^{15,16}. We made use of a set of non-small cell lung carcinoma (NSCLC) samples to predict the pathway status and then sorted according to predicted Ras activity. As shown in Fig. 2c, Ras pathway status very clearly correlates with the histological subtype—most of the adenocarcinoma samples exhibit a high probability of Ras deregulation relative to the squamous cell carcinoma samples. Prediction of the status of the other pathways revealed a less distinct pattern, although each tended to be more active in the squamous cell carcinoma samples (Supplementary Fig. 3). This pattern becomes more evident in the analysis shown in Fig. 3. An examination of Ras mutation identified 11 samples with K-Ras mutations, all confined to the adenocarcinomas (indicated by an asterisk in the figure) (Supplementary Table 2). Overall, 14% of NSCLC tumours and 29% of the adenocarcinomas had K-Ras mutations in codon 12. Because nearly all of the adenocarcinomas exhibited Ras pathway deregulation, it seems that deregulation of the Ras pathway is indeed a characteristic of development of adenocarcinoma of the lung, and that this can occur as a result of Ras mutations as well as following other events that deregulate the pathway.

Whereas the analysis of pathway deregulation as shown in Fig. 2c depicts the status of an individual pathway, the real power in this approach is the ability to identify patterns of pathway deregulation, using hierarchical clustering, much the same as identifying patterns of gene expression. We started with an analysis of the lung cancer samples (Fig. 3a, left panel). This analysis distinguished adenocarcinomas from squamous cell carcinomas, driven in part by the Ras pathway distinction. It is also evident that the tumours predicted as exhibiting relatively low Ras activity are generally predicted at higher levels of Myc, E2F3, β-catenin and Src activity (clusters 1–3). Conversely, the tumours with relatively elevated Ras activity exhibited relatively lower levels of these other pathways (clusters 4–7). Independent of the tumour histopathology, concerted deregulation

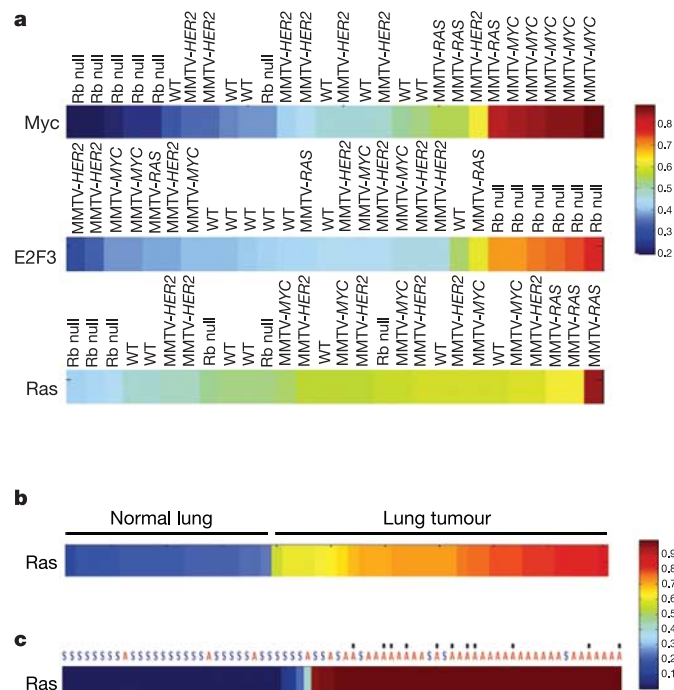


Figure 2 | Validation of pathway predictions in tumours. **a**, Mouse mammary tumours derived from mice transgenic for the MMTV-MYC (five samples), MMTV-HRAS (three samples) or MMTV-HER2 (seven samples) oncogenes, tumours dependent on loss of Rb (six samples), or seven samples of normal mammary tissue were used to verify accuracy and specificity of our signatures. The predicted probability of Myc, E2F3 and Ras activity in mouse tumours was sorted from low (blue) to high (red), and displayed as a colour bar. **b**, Prediction of pathway status in mouse lung cancer model. A set of previously published mouse Affymetrix expression data comparing normal and tumour lung tissue with spontaneous activating KRAS mutations¹⁴ was used to validate the predictive capacity of the Ras pathway signature. The predicted probability of Ras activity in the normal and tumour tissue was sorted from low (blue) to high (red), and displayed as a colour bar. **c**, Relationship of Ras pathway status in NSCLC samples to cell type of tumour origin. The corresponding tumour cell type is indicated as either squamous (S) or adenocarcinoma (A). Ras mutation status is indicated by an asterisk.

of Ras with β -catenin, Src and Myc (cluster 8) identified a population of patients with poor survival—a median survival of 19.7 months versus 51.3 months for all other clusters (Fig. 3a, right panel). Furthermore, this subpopulation of patients exhibited worse survival than any of the groups of patients identified based on the status of any single pathway deregulation (Supplementary Fig. 4). This analysis demonstrates the ability of integrated pathway analysis, based on multiple signatures of component pathway deregulation, to define improved categorization of lung cancer patients.

Two additional examples made use of large sets of breast cancer samples (Fig. 3b) and ovarian cancer samples (Fig. 3c). Again, there were evident patterns of pathway deregulation, distinct from that seen in the lung samples, which characterized the breast and ovarian tumours. For breast cancer, there were two clusters of patients with good prognosis (clusters 2 and 4), and two clusters with poor prognosis (clusters 1 and 3). Furthermore, clusters 2 and 3, which both contain oestrogen receptor (ER)-positive tumours (and no discernable differences in HER2 status or other clinical parameters), show distinct survival rates (P -value = 0.07). Patients defined by cluster 5 (in which higher than average β -catenin and Myc activities were predicted, and E2F3 activity was lower than average) exhibited very poor survival, again illustrating the importance of co-deregulation of multiple oncogenic pathways as a determinant of clinical outcome. A final analysis made use of an advanced stage (III or IV) ovarian cancer data set. The ovarian samples exhibited a dominant pattern of β -catenin and Src deregulation, either elevated (cluster 1 and 2) or diminished (clusters 3–6). Notably, the co-deregulation of Src and β -catenin defined by clusters 1 and 2 identifies

a population of patients with very poor survival compared to other pathway clusters (median survival: 29.0 months versus 91.0 months) (Fig. 3c, right panel). Once again, for these cases, individual pathway status did not stratify patient subgroups as effectively as patterns of multiple pathway deregulation (Supplementary Fig. 4).

Given the capacity of the gene expression signatures to predict deregulation of oncogenic signalling pathways, we have also addressed the extent to which this could predict sensitivity to a therapeutic agent that targets that pathway. To explore this, we predicted pathway deregulation in a series of breast cancer cell lines to be screened against potential therapeutic drugs. The results using the set of five pathway predictors, together with an initial collection of breast cancer cell lines, are shown in Fig. 4a. Biochemical characteristics of the cell lines relevant for pathway analysis are summarized in Supplementary Table 3 and Supplementary Fig. 5. In each case, the relative probabilities of pathway activation are predicted from the signature in a manner completely analogous to the prediction of pathway status in tumours. In most cases, there is a good correlation between biochemical measures of pathway activation and prediction based on gene expression signatures. An exception is with Ras, where there is not a significant correlation between the biochemical measure of pathway activation and pathway prediction, presumably reflecting additional events not measured in the biochemical assay. Clearly, the critical issue is whether the gene expression signature predicts drug sensitivity—this point is addressed by the dose–response assays in Fig. 4b.

In parallel with mapping the pathway status, the cell lines were assayed with drugs known to target specific activities within given

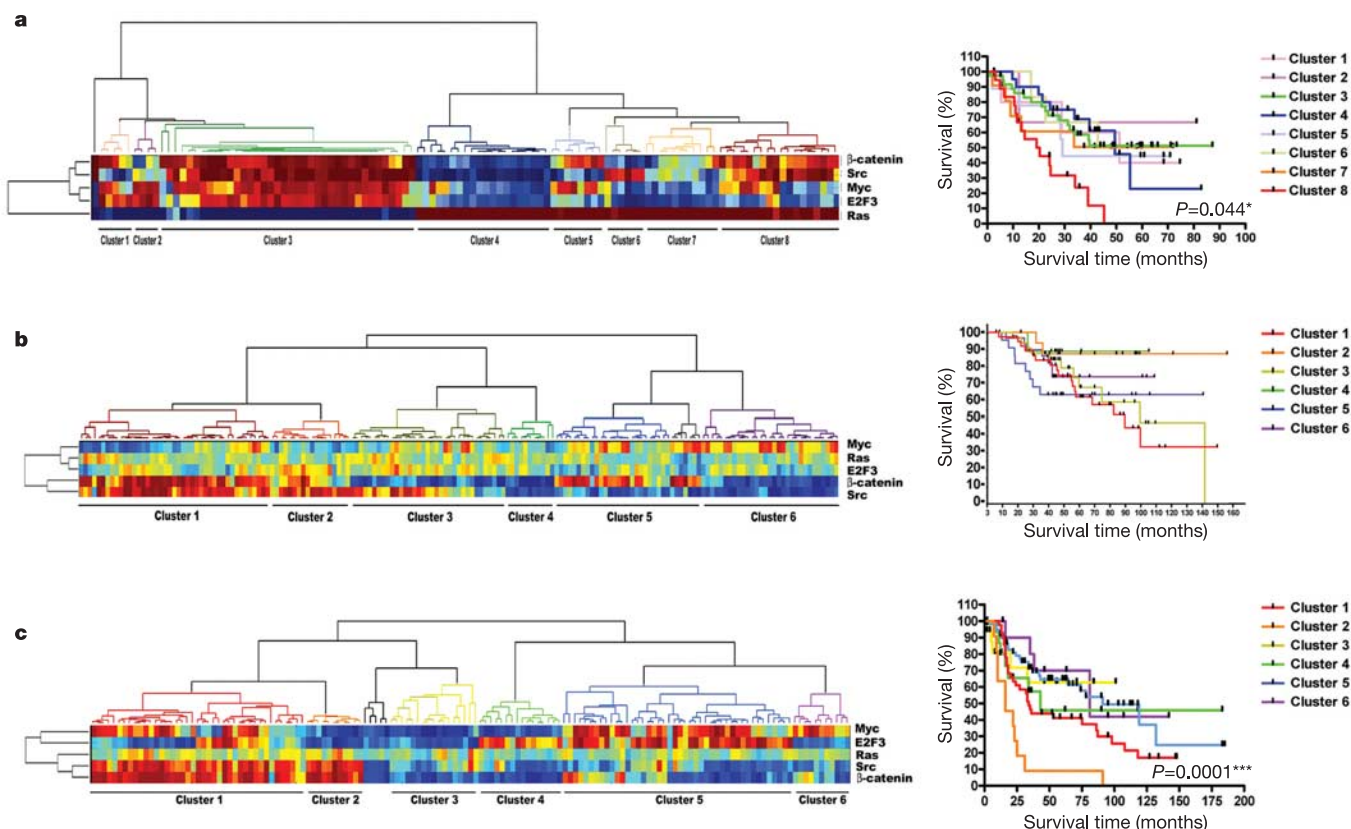


Figure 3 | Patterns of pathway deregulation in human cancers.

a, Hierarchical clustering of predictions of pathway deregulation in samples of human lung tumours (left panel). Prediction of Ras, Myc, E2F3, β -catenin and Src pathway status for each tumour sample was independently determined using supervised binary regression analysis, as described. Red indicates high probability of pathway activation, with blue indicating a low probability. Patterns in the tumour pathway predictions were identified by

hierarchical clustering, and separate clusters are indicated by coloured dendrograms. Kaplan–Meier survival analysis for lung cancer patients based on pathway clusters (right panel). Patient clusters with correlative pathway deregulation shown in the left panel correspond to clusters comprising each independent survival curve. Black tick marks represent censored patients.

b, Breast cancer. Same as for **a**. **c**, Ovarian cancer. Same as for **a**.

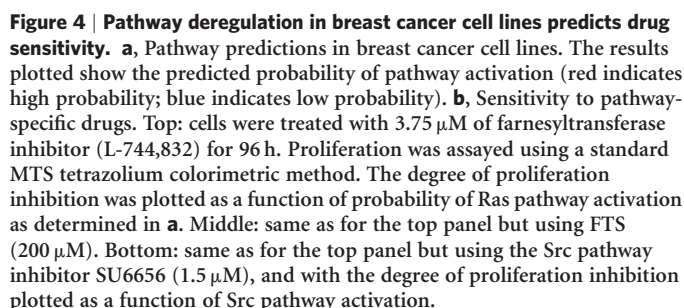
(Fig. 4b). Furthermore, comparison of the drug inhibition results with predictions of other pathways failed to demonstrate a significant correlation (Supplementary Fig. 6). These results confirm the ability of the defined ‘pathway deregulation signatures’ to also predict sensitivity to therapeutic agents that target the corresponding pathways.

Although the development of targeted biological agents holds the promise of a more precise matching of therapy with disease mechanism, it is nevertheless true that the success rate of single agents as well as the selection of combination therapies could be improved. The ability to predict the deregulation of various oncogenic pathways through gene expression analysis offers an opportunity to identify new therapeutic options for patients by providing a potential basis for guiding the use of pathway-specific drugs. The major value of this approach may be the capacity to direct combinations of therapies—multiple drugs that target multiple pathways—based on information that specifies the activation state of the pathways.

Cell and RNA preparation. Human mammary epithelial cells from a breast reduction surgery at Duke University were isolated and cultured according to previously published protocols¹⁹. These cells were a gift from G. Huper. Cells are brought to quiescence, and then infected with adenovirus expressing either human c-Myc, activated H-Ras, human c-Src, human E2F3, or activated β -catenin (we thank J. Kitajewski, W. El-Deiry, Z. German and D. Kuppuswamy for DNA constructs and adenovirus). Eighteen hours after infection, cells were collected and expression of oncogenes and their secondary targets was determined by a standard western blotting protocol (see Supplementary Information). Activation status of kinase pathways for the breast cancer cell lines was determined for growing cells using the following methods. Ras activation is measured using a Ras activation assay kit (Upstate Biotechnology). c-Src activation was determined by western blotting using a phospho-Tyr 416 Src antibody. E2F3, Myc and β -catenin activity was measured by isolating nuclear extracts from cells as previously described, and performing western blotting analysis using antibodies specific for each oncoprotein. Total RNA was extracted for cell lines using the Qiashredder and Qiagen RNeasy Mini kit. Quality of the RNA was checked by an Agilent 2100 Bioanalyser.

DNA microarray analysis. Samples were prepared according to the manufacturer's instructions and as previously published^{21,22}. Experiments to generate signatures use Human U133 2.0 Plus GeneChips. Breast tumours were hybridized to Hu95Av2 arrays, ovarian tumours to Hu133A arrays, and lung tumours to Human U133 2.0 plus arrays (Affymetrix). All microarray data are available at <http://data.cgt.duke.edu/oncogene.php> and on GEO.

Statistical analysis methods. Analysis of expression data is as previously described¹². Briefly, before statistical modelling, gene expression data are filtered to exclude probe sets with signals present at background noise levels, and probe sets that do not vary significantly across samples. Each signature summarizes its constituent genes as a single expression profile, and is here derived as the top principal components of that set of genes. When predicting the pathway activation of cancer cell lines or tumour samples, gene selection and identification is based on the training data, and then metagene values are computed using the principal components of the training data and additional cell line or tumour expression data. Bayesian fitting of binary probit regression models to



the training data then permits an assessment of the relevance of the metagene signatures in within-sample classification, and estimation and uncertainty assessments for the binary regression weights mapping metagenes to probabilities of relative pathway status. Predictions of the relative pathway status of the validation cell lines or tumour samples are then evaluated, producing estimated relative probabilities—and associated measures of uncertainty—of activation/deregulation across the validation samples. Hierarchical clustering of tumour predictions was performed using Gene Cluster 3.0 (ref. 23). Genes and tumours were clustered using average linkage with the uncentred correlation similarity metric. Standard Kaplan–Meier mortality curves and their significance levels were generated for clusters of patients with similar patterns of oncogenic pathway deregulation using GraphPad software. For the Kaplan–Meier survival analyses, the survival curves are compared using the logrank test. For full details, see Supplementary Information.

Cell proliferation assays. Growth curves and dosing ranges for the breast cancer cell lines were carried out as described in Supplementary Information. Sensitivity to a farnesyl transferase inhibitor (L-744,832), FTS and an Src inhibitor (SU6656) was determined by quantifying the percentage reduction in growth (versus DMSO controls) at 96 h using a standard MTS (3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulphophenyl)-2H-tetrazolium) colorimetric assay (Promega). Concentrations used were from 100 nM to 10 μ M L-744,832, 10 to 200 μ M FTS, and 300 nM to 10 μ M SU6656. All experiments were repeated at least three times.

K-Ras mutation assay. K-Ras mutation status was determined using restriction fragment length polymorphism and sequencing as previously described²⁴. See Supplementary Information for additional details.

Received 11 August; accepted 3 October 2005.

Published online 6 November 2005.

1. Fearon, E. R. & Vogelstein, B. A genetic model for colorectal tumorigenesis. *Cell* **17**, 671–674 (1990).
2. Hanahan, D. & Weinberg, R. A. The hallmarks of cancer. *Cell* **100**, 57–70 (2000).
3. Sherr, C. J. Cancer cell cycles. *Science* **274**, 1672–1677 (1996).
4. Ramaswamy, S. & Golub, T. R. DNA microarrays in clinical oncology. *J. Clin. Oncol.* **20**, 1932–1941 (2002).
5. Lamb, J. *et al.* A mechanism of cyclin D1 action encoded in the patterns of gene expression in human cancer. *Cell* **114**, 323–334 (2003).
6. Huang, E. *et al.* Gene expression phenotypic models that predict the activity of oncogenic pathways. *Nature Genet.* **34**, 226–230 (2003).
7. Black, E. P. *et al.* Distinct gene expression phenotypes of cells lacking Rb and Rb family members. *Cancer Res.* **63**, 3716–3723 (2003).
8. Segal, E., Friedman, N., Koller, D. & Regev, A. A module map showing conditional activity of expression modules in cancer. *Nature Genet.* **36**, 1090–1098 (2004).
9. Rhodes, D. R. *et al.* Large-scale meta-analysis of cancer microarray data identifies common transcriptional profiles of neoplastic transformation and progression. *Proc. Natl Acad. Sci. USA* **101**, 9309–9314 (2004).
10. Ramaswamy, S., Ross, K. N., Lander, E. S. & Golub, T. R. A molecular signature of metastasis in primary solid tumors. *Nature Genet.* **33**, 49–54 (2003).
11. Mootha, V. K. *et al.* PGC-1 α -responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nature Genet.* **34**, 267–273 (2003).
12. West, M. *et al.* Predicting the clinical status of human breast cancer by using gene expression profiles. *Proc. Natl Acad. Sci. USA* **98**, 11462–11467 (2001).
13. D'Crus, C. M. *et al.* c-MYC induces mammary tumorigenesis by means of a preferred pathway involving spontaneous Kras2 mutations. *Nature Med.* **7**, 235–239 (2001).
14. Sweet-Cordero, A. *et al.* An oncogenic KRAS2 expression signature identified by cross-species gene expression analysis. *Nature Genet.* **37**, 48–54 (2005).
15. Rodenhuis, S. *et al.* Mutational activation of the K-ras oncogene and the effect of chemotherapy in advanced adenocarcinoma of the lung: a prospective study. *J. Clin. Oncol.* **15**, 285–291 (1997).
16. Salgia, R. & Skarin, A. T. Molecular abnormalities in lung cancer. *J. Clin. Oncol.* **16**, 1207–1217 (1998).
17. Cory, A. H. Use of an aqueous soluble tetrazolium/formazan assay for cell growth assays in culture. *Cancer Commun.* **3**, 207–212 (1991).
18. Riss, T. L. & Moravec, R. A. Comparison of MTT, Xtt, and a novel tetrazolium compound for MTS for *in vitro* proliferation and chemosensitivity assays. *Mol. Biol. Cell* **3**, 184a (1992).
19. Stampfer, M. R. & Yaswen, P. Culture systems for study of human mammary epithelial cell proliferation, differentiation, and transformation. *Cancer Surv.* **18**, 7–34 (1993).
20. Huang, E. *et al.* Gene expression predictors of breast cancer outcomes. *Lancet* **361**, 1590–1596 (2003).
21. Irizarry, R. A. *et al.* Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* (in the press).
22. Bolstad, B. M., Irizarry, R. A., Astrand, M. & Speed, T. P. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19**, 185–193 (2003).
23. Eisen, M. B., Spellman, P. T., Brown, P. O. & Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl Acad. Sci.* **95**, 14863–14868 (1998).
24. Mitsudomi, T. *et al.* Mutations of ras genes distinguish a subset of non-small-cell lung cancer cell lines from small-cell lung cancer cell lines. *Oncogene* **6**, 1353–1362 (1991).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to P. Bild for the inspiration to pursue this research. We also thank K. Shianna, J. Freedman, S. Mori, L. Jakoi and K. Culler for their assistance. A.H.B. has been supported by an AACR-Anna D. Barker Fellowship and an ACS Postdoctoral Fellowship. This work was supported by grants from the NIH (to J.R.N.).

Author Contributions A.H.B. was responsible for all experimental work and computational data analysis. A.H.B. and J.R.N. were responsible for project planning and data analysis. G.Y., J.T.C. and Q.W. were responsible for generation of specialized computer programs used in these studies. H.K.D. and A.P. provided intellectual input and data management support. D.C. and M.-B.J. provided technical support for experiments. M.W. was responsible for conception of the statistical approach and intellectual input. A.B., J.M.L., J.R.M., J.A.O. and D.H. were responsible for the development of clinical resources used in the study.

Author Information The GEO accession numbers for the datasets are: GSE3156, breast cancer cell lines; GSE3158, mouse tumour data set; GSE3151, oncogene signature data set; GSE3141, lung cancer data set; GSE3143, breast cancer data set; GSE3149, ovarian cancer data set. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.R.N. (j.nevins@duke.edu).

LETTERS

BRAF mutation predicts sensitivity to MEK inhibition

David B. Solit^{1,3}, Levi A. Garraway^{4,6}, Christine A. Pratilas^{2,3}, Ayana Sawai³, Gad Getz⁶, Andrea Basso^{3†}, Qing Ye³, Jose M. Lobo³, Yuhong She³, Iman Osman⁷, Todd R. Golub^{5,6}, Judith Sebolt-Leopold⁸, William R. Sellers^{4,6} & Neal Rosen^{1,3}

The kinase pathway comprising RAS, RAF, mitogen-activated protein kinase kinase (MEK) and extracellular signal regulated kinase (ERK) is activated in most human tumours, often through gain-of-function mutations of RAS and RAF family members¹. Using small-molecule inhibitors of MEK and an integrated genetic and pharmacologic analysis, we find that mutation of BRAF is associated with enhanced and selective sensitivity to MEK inhibition when compared to either 'wild-type' cells or cells harbouring a RAS mutation. This MEK dependency was observed in BRAF mutant cells regardless of tissue lineage, and correlated with both downregulation of cyclin D1 protein expression and the induction of G1 arrest. Pharmacological MEK inhibition completely abrogated tumour growth in BRAF mutant xenografts, whereas RAS mutant tumours were only partially inhibited. These data suggest an exquisite dependency on MEK activity in BRAF mutant tumours, and offer a rational therapeutic strategy for this genetically defined tumour subtype.

Activating RAS and BRAF mutations typically demonstrate mutual exclusivity in tumours^{1–3}. This suggests an epistatic relationship whereby either mutation is sufficient to deregulate a common effector pathway such as the MEK–ERK kinase cascade. If so, tumours arising as a result of mutation to either RAS or BRAF should harbour similar downstream dependencies that might represent useful therapeutic targets. To test this hypothesis, we examined the consequences of MEK–ERK pathway inhibition in a collection of cancer cell lines that exhibited differing mechanisms of MAP kinase pathway deregulation. Cell lines containing the NRAS(Q61R) or BRAF(V600E) mutations (present in ~15% and ~50% of melanomas, respectively) were analysed alongside a panel of cancer cell lines that lacked both mutations (hereafter referred to as RAS/BRAF-WT). Several of these RAS/BRAF-WT cell lines exhibit levels of ERK phosphorylation comparable to those observed in the setting of RAS or RAF mutation.

MEK1/2 are dual-specificity kinases that phosphorylate and activate ERK, the classical MAP kinase⁴. To inhibit MEK–ERK, we used the potent and selective MEK inhibitor CI-1040 (ref. 5). CI-1040 is a non-competitive inhibitor of MEK1/2 with a K_i of 300 nM *in vitro*^{5,6}. The only other known CI-1040 target is the MEK5 kinase; however, its inhibition occurs at a 100-fold greater concentration than that required for inhibition of MEK1/2 (ref. 7). Because ERK is the only known MEK substrate, we reasoned that selective MEK inhibition might clarify the role of the MAP kinase pathway in differing genetic contexts.

CI-1040 inhibited MEK (as measured by phosphorylated ERK

(p-ERK) levels) with a half-maximal inhibitory concentration (IC_{50}) of 100–500 nM in all cell lines tested (Fig. 1 and data not shown). In contrast, the IC_{50} for growth inhibition by CI-1040 differed markedly in a manner that correlated with the mechanism of ERK activation (Fig. 1a). Whereas RAS/BRAF-WT cells exhibited resistance to

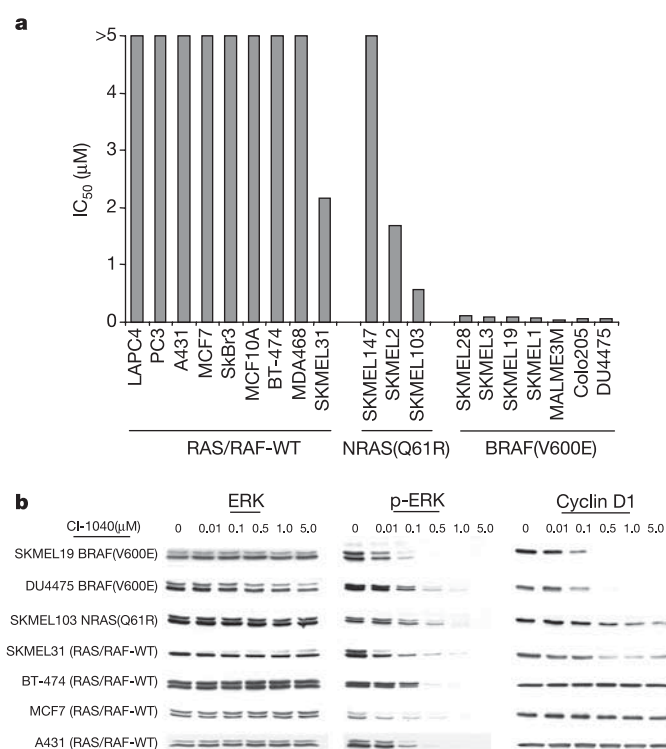


Figure 1 | The BRAF(V600E) mutation confers sensitivity to the MEK inhibitor CI-1040. **a**, CI-1040 IC_{50} values as a function of BRAF and NRAS mutational status. **b**, Immunoblot of p-ERK and total ERK, demonstrating that CI-1040 inhibits MAPK activity with IC_{50} values ranging from 100 to 500 nM. Cells were treated for 24 h. MEK inhibition caused profound downregulation of cyclin D1 expression in BRAF mutant tumour cells. In contrast, cyclin D1 declined only modestly in SKMEL103 melanoma cells with the NRAS(Q61R) mutation and in SKMEL31 RAS/BRAF-WT cells. Cyclin D1 expression was unaffected by CI-1040 in non-melanoma cells with wild-type RAS and BRAF.

¹Department of Medicine, ²Department of Pediatrics, and ³Department of Molecular Pharmacology and Chemistry, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, New York 10021, USA. ⁴Department of Medical Oncology and ⁵Department of Pediatric Oncology, Dana-Farber Cancer Institute, Harvard Medical School, 44 Binney Street, Boston, Massachusetts 02115, USA. ⁶Broad Institute of Harvard and MIT, 320 Charles Street, Cambridge, Massachusetts 02141, USA. ⁷Departments of Medicine and Urology, New York University Medical Center, 550 First Avenue, New York, New York 10016, USA. ⁸Pfizer Global Research and Development, 2800 Plymouth Road, Ann Arbor, Michigan 48105, USA. [†]Present address: Schering-Plough, 2015 Galloping Hill Rd, Kenilworth, New Jersey 07033, USA.

CI-1040 even at concentrations in vast excess of those required for ERK inhibition, cells harbouring a BRAF mutation were exquisitely sensitive, with IC₅₀ values of 0.024–0.111 μM (Fig. 1a). Surprisingly, RAS mutant cells did not demonstrate the same sensitivity despite effective inhibition of p-ERK (Fig. 1b and data not shown). These data raised the possibility that RAS and BRAF mutant cancer cells might be differentially dependent on signalling mechanisms that involve MEK, despite their known epistatic relationship in human cancers.

To explore this hypothesis in an unbiased manner, we interrogated the large-scale chemical sensitivity data available for the NCI60 cancer cell lines⁸ using supervised learning methods previously applied to the analysis of gene-expression data. NCI60 cell lines were partitioned into two classes according to the presence or absence of the BRAF(V600E) mutation. We then performed a supervised analysis⁹ where the mean $-\log_{10}(\text{GI}_{50})$ values for each compound in the BRAF(V600E) and non-mutant classes were compared using a variance fixed *t*-test metric and ranked according to T-score (the GI₅₀ is the concentration that inhibits cell growth by 50%). Thirty-six compounds exhibited significantly increased potency against the BRAF(V600E) class distinction (Fig. 2a and Supplementary Table S1; false discovery rate (FDR) = 0.25, nominal *P* value < 3 × 10^{−4}). The top-scoring compound against the

BRAF(V600E) class was hypothemycin (a resorcylic acid lactone, the homologues of which possess potent and selective MEK inhibitory activity), which was found to inhibit p-ERK at a potency comparable to CI-1040 (Supplementary Fig. S1)^{10,11}. Additional top-scoring compounds included protein LF (anthrax lethal factor), a zinc metalloproteinase known to inactivate MEK through enzymatic cleavage¹², and PD98059 (ref. 13), a well-characterized MEK inhibitor. Thus, at least three of the most potent compounds against the BRAF(V600E) class distinction appeared to exert their effects through MEK inhibition. These results were consistent with the CI-1040 analysis and suggest that BRAF mutation might confer a preferential sensitivity to MEK inhibition in human cancer cells.

NCI60 cell lines that harbour RAS mutations are non-overlapping with respect to the BRAF(V600E) mutation, supporting the notion of a redundant pathway function⁸. To explore this further, supervised analysis of the NCI60 data was repeated, using the class distinction RAS mutant versus wild-type RAS. Surprisingly, and in contrast to the results observed for the BRAF(V600E) class distinction, no compound surpassed the Bonferroni significance threshold in the RAS mutant class (Fig. 2b). Conceivably, RAS and BRAF mutations might elicit similar dependencies despite these results, and our failure to identify compounds in the RAS analysis might reflect confounding genetic heterogeneity. Thus, we performed an additional supervised

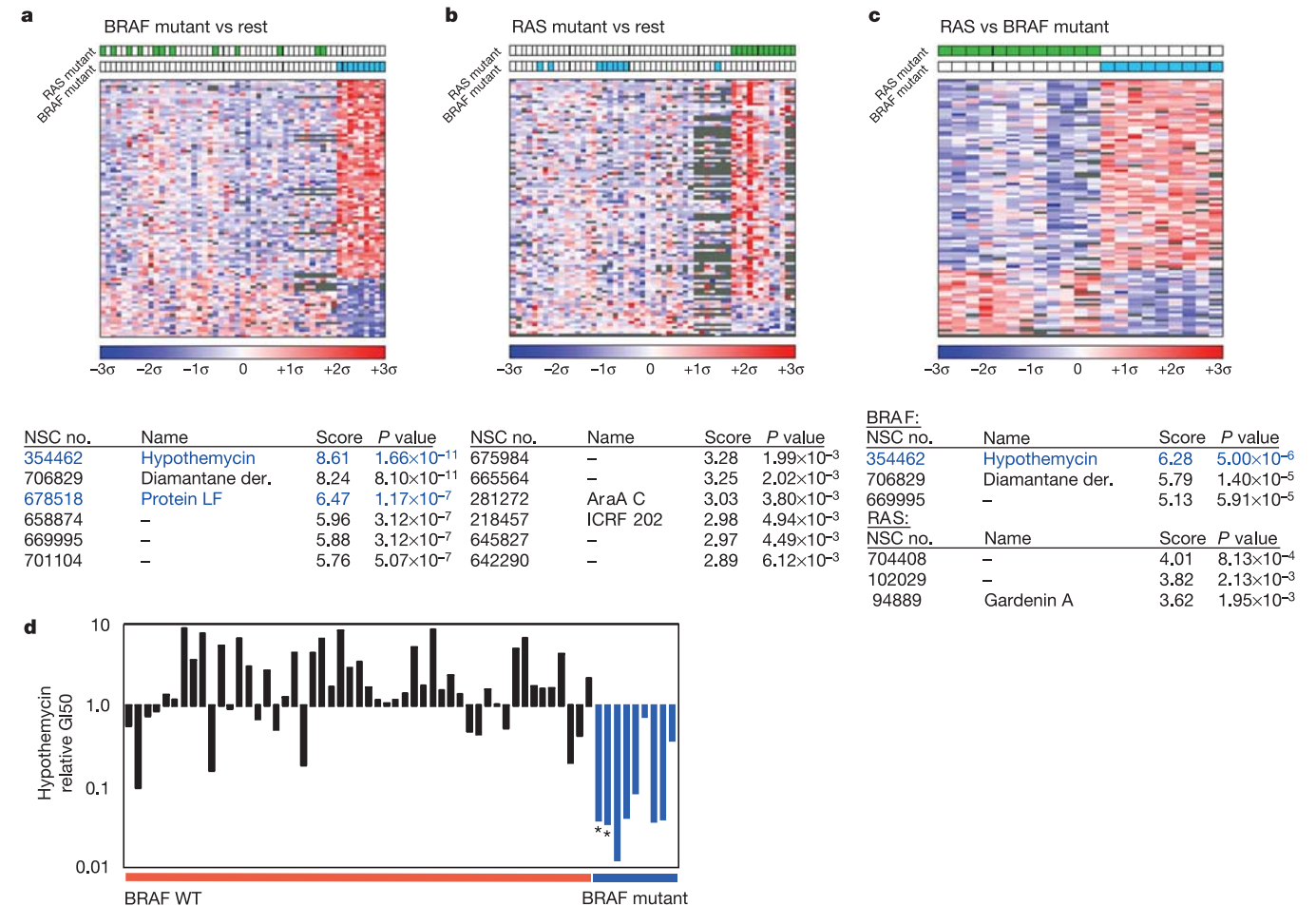


Figure 2 | Chemical sensitivity associated with mutant BRAF and RAS class distinctions. **a–c**, Colourgrams show BRAF mutant (**a**) or RAS mutant (**b**) versus the remaining NCI60 cells, or for mutant RAS versus mutant BRAF lines (**c**). Columns denote NCI60 cell lines; rows denote compounds; colour denotes the number of standard deviations above (red) or below (blue) the mean for all cell lines (top 100 compounds for each class distinction shown; see Methods). NSC numbers, names, variance-fixed T-scores (absolute

values; see Methods) and asymptotic *P* values are shown for top-scoring compounds. Blue font indicates known MEK inhibitors. **d**, Relative GI₅₀ values for the MEK inhibitor hypothemycin in non-haematological NCI60 cell lines. Black bars indicate BRAF wild-type cells; blue bars indicate BRAF(V600E) cells; asterisks indicate non-melanoma cell lines with the BRAF(V600E) mutation.

analysis that directly compared only the BRAF and RAS mutant lines. If these classes do indeed manifest common genetic dependencies, compounds that target the relevant mechanisms (for example, the MEK–ERK pathway) should fail to score by this class distinction. However, hypothemycin again distinguished BRAF and RAS mutant cells (Fig. 2c); protein LF also retained a high rank.

Because most BRAF(V600E) cell lines analysed were melanoma-derived, the enhanced sensitivity to MEK inhibition may have reflected a melanocytic lineage effect independent of the BRAF(V600E) mutation; however, several lines of evidence rendered this possibility unlikely. First, in the NCI60 analyses all BRAF(V600E) cell lines exhibited markedly reduced hypothemycin GI50 values relative to the mean across the sample set, regardless of tissue type (Fig. 2d). Colo205, an NCI60 colon cancer line with the BRAF(V600E) mutation, was also found to exhibit sensitivity to CI-1040 at an equivalent level to the melanoma cells (Fig. 1a). In addition, the two melanoma lines lacking a BRAF mutation were clearly indifferent to the effects of hypothemycin. Finally, only one of the breast/prostate cell lines demonstrated similar sensitivity to the drug: the breast cancer line DU-4475 (IC₅₀ 24 nM). Notably, sequencing of BRAF in this cell line showed that it also contained a V600E mutation. Thus, the sensitivity of cancer cell lines to MEK inhibition correlated most closely with BRAF mutation status.

In many cell types, RAS–RAF–MEK–ERK signalling is required for both D-cyclin expression and assembly of the cyclin D–cdk4 complex¹⁴. The marked sensitivity of BRAF mutant cells to MEK inhibitors allowed us to examine the functional consequences of MAP kinase blockade in this context. Treatment of BRAF mutant cell lines with CI-1040 caused a marked decline in D-cyclin protein levels (Figs 1b and 3a, d). In the SKMEL28 cell line, this decline was followed by loss of RB phosphorylation and a profound G1 cell cycle arrest (Fig. 3). G1 arrest was accompanied by apoptosis in several BRAF mutant cell lines (Fig. 3c, d), suggesting that MEK inhibition in a BRAF-mutant context exerts both cytotoxic and cytostatic effects.

In contrast, CI-1040 concentrations that completely inhibited p-ERK had no effect on cyclin D1 protein expression in the vast majority of RAS/BRAF-WT cells, as shown in Fig. 1b for the MCF7, BT-474 and A431 cell lines. BT-474 and A431 exhibited robust MAPK activity driven by HER2/neu and EGFR, respectively, suggesting that cyclin D1 expression and G1 progression are driven by MEK/ERK-independent mechanisms in certain RAS/BRAF-WT cells.

To determine whether the differential sensitivity to MEK inhibition observed for BRAF mutant cancer cells was re-capitulated *in vivo*, mice harbouring xenograft tumours were treated with the MEK inhibitor PD0325901. PD0325901 is a derivative of CI-1040 that has improved oral bioavailability and induces a longer duration of target suppression¹⁵. The effects of PD0325901 on tumour cells *in vitro* are qualitatively identical to those of CI-1040, including the marked selectivity for BRAF mutant cell lines, but occur at 100-fold lower concentrations (Supplementary Fig. S2).

Daily treatment with PD0325901 at doses of 5 and 25 mg kg⁻¹ completely suppressed the growth of SKMEL28 and Colo205 BRAF(V600E) mutant xenografts (Fig. 4a and Supplementary Fig. S3; $P < 0.01$ for both 5 and 25 mg kg⁻¹ versus control, $P = 0.16$ for 5 versus 25 mg kg⁻¹). Growth suppression was associated with loss of D-cyclin expression, induction of p27 and hypophosphorylation of RB (Fig. 4e and Supplementary Fig. S4). In contrast, PD0325901 treatment of SKMEL103 (NRAS(Q61R)), SKMEL30 (NRAS(Q61R)) and SKMEL31 (RAS/BRAF-WT) xenografts at a dose of 5 mg kg⁻¹ only delayed tumour growth, with complete growth suppression requiring 25 mg kg⁻¹ (Fig. 4b, c and Supplementary Fig. S3b; $P < 0.01$ for 5 versus 25 mg kg⁻¹, and 5 and 25 mg kg⁻¹ versus control). BT-474 xenografts (BRAF/RAS-WT) were completely insensitive to PD0325901 (Fig. 4d). PD0325901 treatment at the doses studied was non-toxic and resulted in profound p-ERK inhibition in all xenograft models studied; however, RB phosphorylation, cyclin D expression and proliferation as measured by Ki67 were unaffected by MEK inhibition in

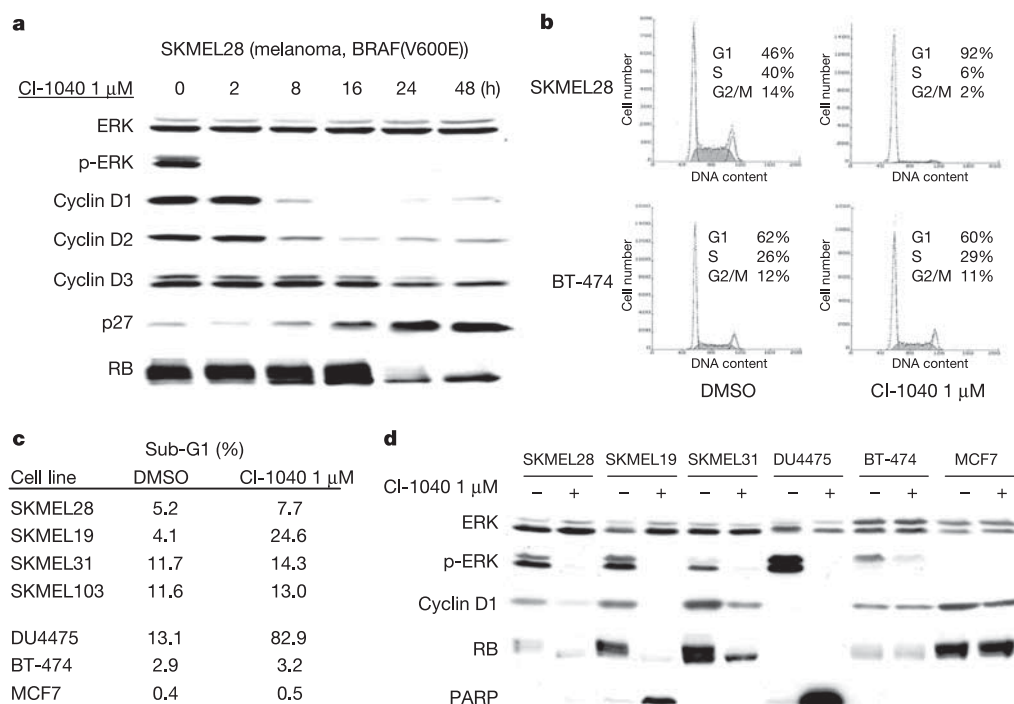


Figure 3 | MEK inhibition causes loss of D-cyclin expression, RB hypophosphorylation and G1 arrest in BRAF mutant cancer cells.

a, Immunoblot showing the kinetics of change in p-ERK, D-cyclin expression and RB in SKMEL28 cells treated with 1 μM CI-1040. **b**, CI-1040

induced a G1 growth arrest in BRAF mutant tumour cells but not in RAS/BRAF-WT breast cancer cells (BT-474 shown). **c**, **d**, CI-1040 induces apoptosis in some but not all cancer cell lines with the BRAF(V600E) mutation as measured by FACS analysis (**c**) and PARP cleavage (**d**).

PD0325901-resistant BT-474 xenografts (Fig. 4e, f), and there was no correlation between basal p-ERK levels and PD0325901 sensitivity (Supplementary Fig. S4). Thus, the MEK dependency characteristic of BRAF mutant tumour cells *in vitro* was also apparent *in vivo*.

Excess MAP kinase pathway activation occurs commonly in human tumours. In melanoma and other solid tumours, mutation of BRAF and RAS occurs frequently and tends to exhibit mutual exclusivity, suggesting that each mutation confers a similar selective advantage¹. However, our findings suggest that tumour cells carrying BRAF mutations are much more reliant on MEK–ERK signalling than are RAS mutant cells, or cells that activate MAP kinase by other means. Thus, BRAF mutant cancer cells may harbour a critical dependency on MEK–ERK that renders them highly sensitive to pharmacological MEK inhibition.

BRAF mutations occur at a high frequency in melanomas, but are also observed in colon, lung and several other tumour types^{1,2}. Expression of BRAF(V600E) in non-transformed melanocytes leads to constitutive ERK activation and tumorigenicity in mice, and depletion of BRAF but not A-RAF or C-RAF in BRAF(V600E) mutant melanoma cells reduces ERK activity^{16,17}. Our data suggest

that D-cyclin expression is also deregulated and ERK-dependent in BRAF-mutant tumours. Cyclin D downregulation may therefore mediate at least some of the anti-proliferative effects observed after MEK inhibition. On the other hand, MEK inhibition had little effect on D-cyclin expression in most BRAF/RAS-WT tumour cells. In these cells, mutations in the PTEN or phosphatidylinositol-3-OH kinase (PI(3)K) genes, or activation of other pathways, may drive D-cyclin expression in an ERK-independent fashion^{18,19}. Our results are also consistent with a model in which ERK regulates G1 progression only in certain lineages (for example, melanocytes); presumably, such lineage differences in cell growth control contribute to the imbalanced frequency of RAS and BRAF mutations observed across tumour types.

RAS-dependent transformation has been found previously to require activation of cyclin D1 (refs 20–23). As both oncogenic RAF and activated ERK also induce cyclin D1 expression^{24,25}, it has been presumed that in human tumours with RAS mutation, cyclin D1 expression was controlled by RAS-mediated MEK–ERK activation. However, our results suggest that in certain genetic contexts, including some tumours with RAS mutation, ERK signalling may be dispensable for cyclin D1 expression and cell proliferation. RAS family members have multiple other targets, such as PI(3)K and RalGDS; these may exert more prominent oncogenic effects in certain tumour subtypes, thereby reducing the requirement for MEK–ERK activation^{26,27}. Our findings therefore raise the possibility that single-agent therapeutic strategies may prove insufficient in RAS mutant tumours. Instead, direct RAS inhibitors or combinatorial strategies may be required.

Thus far, the use of BRAF inhibitors in clinical trials has met with mixed results. On the other hand, the favourable therapeutic index and selectivity of MEK inhibitors may provide an appealing therapeutic strategy for BRAF mutant cancers. We therefore propose clinical trials of MEK inhibitors in which patients are stratified based on BRAF mutational status.

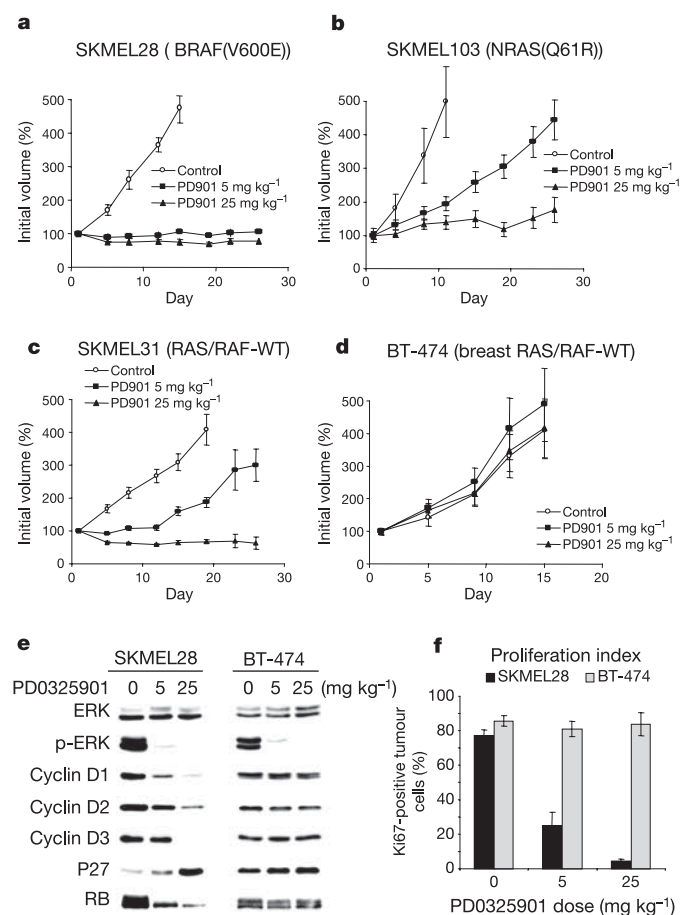


Figure 4 | PD0325901 completely suppresses the growth of BRAF(V600E) mutant xenografts. **a**, PD0325901 suppressed the growth of SKMEL28 (BRAF(V600E)) xenografts at both the 5 and 25 mg kg⁻¹ dose levels. **b**, **c**, In contrast, 5 mg kg⁻¹ PD0325901 only delayed the growth of SKMEL103 (RAS(Q61R)) and SKMEL31 (RAS/RAF-WT) xenografts, with complete growth suppression requiring the higher dose. **d**, BT-474 xenografts (RAS/RAF-WT) were refractory to MEK inhibition (*n* = 10 mice per group). **e**, **f**, PD0325901 reduced p-ERK levels in both SKMEL28 and BT-474 xenograft tumours but caused downregulation of D-cyclins, induction of p27 and hypophosphorylation of RB, and a decline in the proliferative index only in the SKMEL28 xenografts. Tumour lysates were derived from mice euthanized 8 h after the final treatment. All error bars indicate standard error.

METHODS

Cell culture. CI-1040 and PD0325901 were obtained from Pfizer Global Research and Development. Drugs were dissolved in DMSO to yield 10 mM stock solutions and stored at -20 °C. All SKMEL lines were obtained from A. Houghton and P. Chapman with the remainder obtained from the ATCC. Cells were maintained in either RPMI or a 1:1 mixture of DMEM:F12 medium supplemented with 2 mM glutamine, 50 U ml⁻¹ each of penicillin and streptomycin, and 10% heat-inactivated fetal bovine serum, and incubated at 37 °C in 5% CO₂.

Alamar blue cell proliferation assay. Cells were plated in 96-well plates at a density of 2,000–5,000 cells per well. After 24 h, cells were treated with a range of drug concentrations prepared by serial dilution. The cells were exposed to Alamar blue (AccuMed International, OH) 3–5 days after drug treatment, and plates were read using a fluorescence spectrophotometer.

Western blot analysis. Treated cells were harvested, washed with PBS and lysed in NP40 lysis buffer (50 mM Tris (pH 7.4), 1% NP40, 150 mM NaCl, 40 mM NaF, 1 mM Na₃VO₄, 1 mM phenylmethylsulphonylfluoride, and 10 µg ml⁻¹ each of leupeptin, aprotinin and soybean trypsin inhibitor) for 30 min on ice. Lysates were centrifuged at 13,200 r.p.m. for 10 min and the protein concentration of the supernatant was determined by BCA assay (Pierce). Equal amounts of total protein were resolved by SDS-PAGE and transferred onto nitrocellulose membranes. Blots were probed overnight at 4 °C with antibody raised against the protein of interest. Anti-MAP kinase, phospho-MAP kinase, RB and cleaved PARP antibodies were obtained from Cell Signaling Technology. Anti-cyclin D1, anti-cyclin D2 and anti-cyclin D3, and p27 antibodies, were obtained from Santa Cruz Biotechnology. After incubation with horseradish peroxidase-conjugated secondary antibodies, proteins were detected using chemiluminescence (Amersham).

Apoptosis. After drug treatment, both adherent and floating cells were harvested and stained with ethidium bromide. Detection and quantification of apoptotic cells (sub-G1) were performed by flow cytometric analysis.

Animal studies. Four- to six-week-old *nu/nu* athymic female mice were obtained from the National Cancer Institute, Frederick Cancer Center and maintained in ventilated caging. Experiments were carried out under an IACUC approved protocol and institutional guidelines for the proper and humane use of

animals in research were followed. Tumours were generated by injecting $0.5\text{--}1.0 \times 10^7$ tumour cells together with reconstituted basement membrane (Matrigel, Collaborative Research). For the BT-474 model, before tumour cell inoculation, $0.72 \text{ mg pellet}^{-1}$ 17β -estradiol pellets (Innovative Research of America) were inserted subcutaneously. Before initiation of treatment, mice were randomized to receive PD0325901 at a dose of 5 and 25 mg kg^{-1} or vehicle only as control. PD0325901 was formulated in 0.5% hydroxypropyl methylcellulose plus 0.2% Tween 80, and administered by oral gavage. Mice were killed by CO_2 euthanasia. The average tumour diameter (two perpendicular axes of the tumour were measured) was measured in control and treated groups using a calliper. The data are expressed as the increase or decrease in tumour volume in mm^3 ($\text{mm}^3 = \pi/6 \times (\text{larger diameter} \times (\text{smaller diameter})^2)$). Treatment arms were compared using the Wilcoxon rank sums test. To prepare lysates, tumour tissue was homogenized in 2% SDS lysis buffer and then processed as described above. For immunohistochemical studies, xenograft tumours were fixed overnight in paraformaldehyde followed by dehydration in graded ethanols.

Statistical methods. Pharmacological data ($-\log_{10}(\text{GI50})$) for 42,796 compounds were downloaded from the NCI website (http://dtp.nci.nih.gov/docs/cancer/cancer_data.html). The GI50 data were used to populate a matrix with MATLAB software as previously described²⁸. Briefly, where multiple NCS entries existed, the entry with the largest number of replicates was included; in cases where multiple entries had the same number of replicates, the largest mean ($-\log_{10}(\text{GI50})$) value for the NCI60 data set was selected. Incomplete data were assigned as NaN (not a number) for statistical purposes. These GI50 data from all solid tumour NCI60 cell lines were used in a supervised analysis according to the class distinctions described. BRAF mutant status was determined based on published data¹ and by genotyping assays performed by our group²⁸. Because the GI50 data are non-gaussian with many ($-\log_{10}(\text{GI50})$) values at or near 4, a variance-fixed *t*-test was used to calculate significance. Here, the mean and median standard deviation was calculated for compounds for which the mean $-\log_{10}(\text{GI50})$ values across the NCI60 set were between 6 and 7. For both calculations, the standard deviation was near 0.4; thus, this value was used as a minimum threshold standard deviation for the supervised analysis. Compounds with the top variance-fixed T-scores for the relevant class distinctions were selected for additional analysis; in Fig. 2, the absolute values of these scores are indicated. GI50 values and distributions for selected compounds were analysed through the NCI Developmental Therapeutic website.

Received 10 May; accepted 4 October 2005.

Published online 6 November 2005.

- Davies, H. *et al.* Mutations of the *BRAF* gene in human cancer. *Nature* **417**, 949–954 (2002).
- Brose, M. S. *et al.* BRAF and RAS mutations in human lung cancer and melanoma. *Cancer Res.* **62**, 6997–7000 (2002).
- Gorden, A. *et al.* Analysis of BRAF and N-RAS mutations in metastatic melanoma tissues. *Cancer Res.* **63**, 3955–3957 (2003).
- Crews, C. M., Alessandrini, A. & Erikson, R. L. The primary structure of MEK, a protein kinase that phosphorylates the ERK gene product. *Science* **258**, 478–480 (1992).
- Sebolt-Leopold, J. S. *et al.* Blockade of the MAP kinase pathway suppresses growth of colon tumors *in vivo*. *Nature Med.* **5**, 810–816 (1999).
- Ohren, J. F. *et al.* Structures of human MAP kinase kinase 1 (MEK1) and MEK2 describe novel noncompetitive kinase inhibition. *Nature Struct. Mol. Biol.* **11**, 1192–1197 (2004).
- Mody, N., Leitch, J., Armstrong, C., Dixon, J. & Cohen, P. Effects of MAP kinase cascade inhibitors on the MKK5/ERK5 pathway. *FEBS Lett.* **502**, 21–24 (2001).
- Stinson, S. F. *et al.* Morphological and immunocytochemical characteristics of human tumour cell lines for use in a disease-oriented anticancer drug screen. *Anticancer Res.* **12**, 1035–1053 (1992).
- Golub, T. R. *et al.* Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* **286**, 531–537 (1999).
- Zhao, A. *et al.* Resorcylic acid lactones: naturally occurring potent and selective inhibitors of MEK. *J. Antibiot. (Tokyo)* **52**, 1086–1094 (1999).
- Dombrowski, A. *et al.* Production of a family of kinase-inhibiting lactones from fungal fermentations. *J. Antibiot. (Tokyo)* **52**, 1077–1085 (1999).
- Chopra, A. P., Boone, S. A., Liang, X. & Duesbery, N. S. Anthrax lethal factor proteolysis and inactivation of MAPK kinase. *J. Biol. Chem.* **278**, 9402–9406 (2003).
- Alessi, D. R., Cuenda, A., Cohen, P., Dudley, D. T. & Saltiel, A. R. PD 098059 is a specific inhibitor of the activation of mitogen-activated protein kinase kinase *in vitro* and *in vivo*. *J. Biol. Chem.* **270**, 27489–27494 (1995).
- Cheng, M., Sexl, V., Sherr, C. J. & Roussel, M. F. Assembly of cyclin D-dependent kinase and titration of p27Kip1 regulated by mitogen-activated protein kinase kinase (MEK1). *Proc. Natl Acad. Sci. USA* **95**, 1091–1096 (1998).
- Sebolt-Leopold, J. S. & Herrera, R. Targeting the mitogen-activated protein kinase cascade to treat cancer. *Nature Rev. Cancer* **4**, 937–947 (2004).
- Wellbrock, C. *et al.* V599EB-RAF is an oncogene in melanocytes. *Cancer Res.* **64**, 2338–2342 (2004).
- Karasarides, M. *et al.* B-RAF is a therapeutic target in melanoma. *Oncogene* **23**, 6292–6298 (2004).
- Muise-Helmericks, R. C. *et al.* Cyclin D expression is controlled post-transcriptionally via a phosphatidylinositol 3-kinase/Akt-dependent pathway. *J. Biol. Chem.* **273**, 29864–29872 (1998).
- Diehl, J. A., Cheng, M., Roussel, M. F. & Sherr, C. J. Glycogen synthase kinase-3 β regulates cyclin D1 proteolysis and subcellular localization. *Genes Dev.* **12**, 3499–3511 (1998).
- Filmus, J. *et al.* Induction of cyclin D1 overexpression by activated ras. *Oncogene* **9**, 3627–3633 (1994).
- Liu, J. J. *et al.* Ras transformation results in an elevated level of cyclin D1 and acceleration of G1 progression in NIH 3T3 cells. *Mol. Cell. Biol.* **15**, 3654–3663 (1995).
- Albanese, C. *et al.* Transforming p21ras mutants and c-Ets-2 activate the cyclin D1 promoter through distinguishable regions. *J. Biol. Chem.* **270**, 23589–23597 (1995).
- Aktas, H., Cai, H. & Cooper, G. M. Ras links growth factor signalling to the cell cycle machinery via regulation of cyclin D1 and the Cdk inhibitor p27KIP1. *Mol. Cell. Biol.* **17**, 3850–3857 (1997).
- Kerkhoff, E. & Rapp, U. R. Induction of cell proliferation in quiescent NIH 3T3 cells by oncogenic c-Raf-1. *Mol. Cell. Biol.* **17**, 2576–2586 (1997).
- Weber, J. D., Raben, D. M., Phillips, P. J. & Baldassare, J. J. Sustained activation of extracellular-signal-regulated kinase 1 (ERK1) is required for the continued expression of cyclin D1 in G1 phase. *Biochem. J.* **326**, 61–68 (1997).
- Hamad, N. M. *et al.* Distinct requirements for Ras oncogenesis in human versus mouse cells. *Genes Dev.* **16**, 2045–2057 (2002).
- Gonzalez-Garcia, A. *et al.* RalGDS is required for tumour formation in a model of skin carcinogenesis. *Cancer Cell* **7**, 219–226 (2005).
- Garraway, L. A. *et al.* Integrative genomic analyses identify MITF as a lineage survival oncogene amplified in malignant melanoma. *Nature* **436**, 117–122 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The authors thank H. Ju, W. L. Wong and H. Tseng for technical assistance. This work was supported by grants from the National Institutes of Health (L.A.G., C.A.P., G.G., T.R.G., W.R.S. and N.R.), the William H. Goodwin and Alice Goodwin Foundation for Cancer Research, the MSKCC Experimental Therapeutics Program (D.B.S. and N.R.), the Waxman Foundation (D.B.S. and N.R.), the Howard Hughes Medical Institute (G.G. and T.R.G.), Golfers Against Cancer (D.B.S. and N.R.) and the American Society of Clinical Oncology (D.B.S. and C.A.P.).

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare competing financial interests: details accompany the paper at www.nature.com. Correspondence and requests for materials should be addressed to N.R. (rosenn@mskcc.org).

-  FOCUS
-  SPOTLIGHT
-  RECRUITMENT
-  ANNOUNCEMENTS
-  EVENTS

naturejobs

Speaking in tongues

The recent case of a Chinese graduate student who was almost expelled from Yale University and sent home despite passing her qualifying exams (see *Nature* **438**, 278-279; 2005) raises some difficult questions. How much responsibility should foreign students shoulder to ensure that they are prepared for work in their temporary home? And to what extent should the host institution put itself out to ensure their guests' success?

Academic credentials and scientific skills are clearly prerequisites for visiting students, but what about language skills? To find out more, *Naturejobs* conducted a swift survey of its readers. The vast majority of respondents (79%) said the onus should be on the students to improve their language skills, with 39% saying that the visitors should take an intensive language course before heading abroad. Only 11% believed that foreign graduate students should ask their adviser or their peers at the host institution for help overcoming the language barrier.

At first glance, it seems reasonable to expect incoming students to bone up on the language they will be working in before they arrive. But the question really centres on the

social contract between an institution and its students. In many instances, universities not only provide education for foreign graduates, they also benefit from the relatively cheap research and teaching support these students provide. In addition, institutions in the United States and Europe are increasingly viewing foreign scientists as a key component in their success — so it seems only fair for the students to get help acclimatizing in return.

Some places already take this approach in terms of scientific acumen. The European Molecular Biology Laboratory in Heidelberg, Germany, for instance, attracts students from many different education systems, and works to bring everyone to the same scientific level within their first year. Maybe institutions should now be doing the same sort of thing for language and teaching skills.



Paul Smaglik, *Naturejobs* editor

CONTACTS

Publisher: Ben Crowe
Editor: Paul Smaglik
Assistant Editor: Corie Lok

European Head Office, London
 The Macmillan Building, 4 Crinan Street
 London N1 9XW, UK
 Tel: +44 (0) 20 7843 4961
 Fax: +44 (0) 20 7843 4996
 e-mail: naturejobs@nature.com

Naturejobs Sales Director:
 Nevin Bayoumi (4978)
European Sales Manager:
 Andy Douglas (4975)

Natureevents: Sille Opstrup (4994)
UK/RoW/Ireland/Italy:
 Nils Moeller (4953)
 Irene Viglia-Atton (4944)
Scandinavia/Spain/Portugal:
 Evelina Rubio Håkansson (4973)
France/Switzerland/Belgium:
 Amelie Pequignot (4974)
Germany/Austria/The Netherlands:
 Reya Silao (4970)

Advertising Production Manager:
 Billie Franklin
 To send materials use London
 address above.
 Tel: +44 (0) 20 7843 4814

Fax: +44 (0) 20 7843 4996
 e-mail: naturejobs@nature.com
Naturejobs web development: Tom Hancock
Naturejobs online production: Niamh Shields

European Satellite Office
Germany/Austria/
The Netherlands:
 Patrick Phelan
 Tel: +49 89 54 90 57 11
 Fax: +49 89 54 90 57 20
 e-mail: p.phelan@nature.com

US Head Office, New York
 75 Varick Street,
 9th Floor, New York,

NY 10013-1917
 Tel: +1 800 989 7718
 Fax: +1 800 989 7103
 e-mail: naturejobs@natureny.com

US Sales Manager: Peter Bless

Japan Head Office, Tokyo
 Chiyoda Building,
 2-37 Ichigayatamachi,
 Shinjuku-ku,
 Tokyo 162-0843
 Tel: +81 3 3267 8751
 Fax: +81 3 3267 8746
Asia-Pacific Sales Director: Rinoko Asami
 e-mail: rasami@naturejpn.com

Dream big

Being an astronaut, video-game designer or museum curator may be every child scientist's dream. **Kendall Powell** talks to the creative scientists who followed the fantasy.



IMAGES.COM/CORBIS

An Earth scientist, Kathy Sullivan spends many hours contemplating the world around her and how its systems work. But unlike most Earthbound scientists, she has also studied the planet from a space-shuttle window.

In an attempt to sate her curiosity, Sullivan has held several 'dream' jobs in one lifetime: astronaut, Naval reserve captain, presidential appointee and science adviser at the Center of Science and Industry museum in Columbus, Ohio. For her, going from field expeditions to space exploration seemed like a natural progression, though she realizes that hopping through wildly different career options is not for everyone. "It has never been about titles or pay cheques," she says. For Sullivan, it's about discovery and sharing her story.

Scientists who tap into the artistic and adventurous sides of their personalities, like Sullivan, have great stories to share. They use the same problem-solving and technical expertise required for any scientific work, along with extras such as a creative urge, an appreciation of aesthetics or a desire to whet the public's curiosity. On these pages, some of those who have found success in the aerospace, museum and entertainment arenas provide tips for anyone seeking their own dream job.

Rocket science

Astronaut-scientists tend to be generalists and self-starters, like Loren Acton, a solar physicist at Montana State University in Bozeman, who flew into space in 1985 as a payload specialist. He oversaw experiments in solar energy, biology and cosmology. Acton says his experience in flying instruments on rockets and his ability to work to deadlines landed him the job. Even so, he admits the pressures of space science got to him.

"I overdosed on responsibility," he says. In an intense week, he agonized about a couple of blunders that wrecked experiments. In space, there are no second chances — a fact that can catch researchers unaware.

Acton and Sullivan note that chances for scientists to enter the astronaut corps in mid-career, as they did,

have fallen to almost zero, especially given the uncertainty surrounding the shuttle programme. But these lessons also apply to scientists sending research into space from the ground, as Acton still does with rocket-based instruments and satellites.

First Lieutenant Kenyatta Ruffin hopes to become a pilot astronaut, following a more traditional path from Air Force fighter pilot to test pilot and finally to NASA. A degree in aeronautical engineering was the first step towards being entrusted with a \$25-million aircraft.

Ruffin says that flying F-16 Falcons requires as much maths as machismo. He has to calculate those dreaded word problems — if a plane on your radar is 40 miles away and travelling 5 miles a minute, how many minutes before you are in critical range? — in his head. The job also requires a working knowledge of your aircraft and weapons systems as well as those of the enemy.

Problem-solving in his "office at 20,000 feet going at the speed of sound" differs from other scientific disciplines, Ruffin says. But he enjoys approaching problems from many equally acceptable directions. After all, an aerospace-engineering degree could also land you in a job like Christine Lear's, designing Formula 1 race cars for the Sauber-Petronas team in Basel, Switzerland (see *Nature* **431**, 736–738; 2004).

And all high-flying scientists must possess one trait: courage under fire. Sullivan asks: "When things start getting scary or ugly, how do you respond?"

Entertain me

If things turn ugly where Hank Driskill works, he simply sends everyone back to the drawing boards, at Walt Disney Feature Animation in Burbank, California. Like other scientists in the world of Hollywood entertainment, Driskill had an early love of both science and art.

He was inspired in high school by video-arcade games and the movie *Tron*, which featured a hero trapped inside a computer system. He thought how someone must create those digitized worlds, and says



Space stars: Kathy Sullivan and Kenyatta Ruffin.



NASA

M. SULLIVAN/WIREIMAGE.COM

"I realized these people were using math and science to do art". While pursuing a doctorate in computer science, Driskill attended the large computer-graphics conference SIGGRAPH as a student volunteer, aiming to meet people working at visual-effects studios.

Projects he has worked on include designing software tools to simulate the muscle and skin of a Godzilla-like character, dragon fireballs in the live-action movie *Reign of Fire*, and billows of smoke and spray in *Apollo 13*. He draws heavily on his physics background to evoke realistic simulations of fire, fluid and cloth dynamics. He notes that the science and art don't always mesh perfectly when eye-pleasing entertainment is the bottom line.

One might imagine Disney's computer scientists zooming around on scooters wearing mouse ears, but, Driskill says, the fun environment is also a serious one, aiming to produce beautiful and lasting films.

Dan Spitzley, a video-game programmer with Obsidian Entertainment in Santa Ana, says he and his colleagues do a fair amount of game-playing, but also put in long hours to get games on the shelves.

Obsidian focuses on role-playing games such as the forthcoming *Neverwinter Nights 2*, which call for the creation of complex character-player interactions, fantasy worlds and reward systems for advancement. "It's an interesting beast that requires knowledge of many different programming disciplines — graphics, databases, compilers, dialogue tools," says Spitzley, contrasting it with business-software design.

These computer scientists consider themselves entertainers and say that an eye for aesthetics and a love of the product is a must for success.

"If you don't play games and understand what it is about them that gamers expect, your products will reflect that," says Spitzley. Both he and Driskill encourage newcomers to read up on industry magazines or interview industry leaders to understand the business. And play around with available software tools to create a demo game or visual effect.

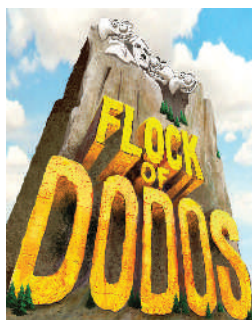
Marine biologist Randy Olson — a writer-director with a PhD in biology from Harvard — has similar advice for scientists who might want to leap into film-making, as he did a decade ago.

"Just start shooting film," he says. With current technology, anyone can become an amateur director. Olson left a tenured position

at the University of New Hampshire in 1994 to attend film school at the University of Southern California.

He has tried screenwriting and directs the Shifting Baselines public awareness campaign on ocean decline, rubbing elbows with celebrity conservationists. Most recently, he has just finished *Flock of Dodos: The Evolution-Intelligent Design Circus*, which he describes as a "polite Michael Moore film". Olson travelled to his home state of Kansas and other parts of the country to challenge proponents of intelligent design.

"It's been almost 15 years for me and I'm still not a billionaire," Olson jokes. "But it's extremely gratifying and fun." Olson sees himself as a communicator



The big picture: top, Randy Olson between actors Ben Stiller and Dustin Hoffman, with conservationist Scott Burns, far right; above, the artwork for Olson's forthcoming film.



Cutting edge: Dan Spitzley and Obsidian's *Knights of the Old Republic II*.

conveying serious scientific messages through humour and popular media.

A Hollywood career requires working and communicating with people "whose brains work differently from yours", says Driskill. Competition among players in this world also has a different spin from that of academia. "It's a

great friendly sense of one-upmanship. We want people to keep upping the artistic bar," Driskill says.

Gee whiz

The artistic bar is considerably lower for the crayon art at the Butterfly Pavilion in Westminster, Colorado, but the environment is no less enchanting. Entomologist Mary Ann Colley is the acting curator at the only free-standing insect zoo in the United States. In the middle of a Colorado winter, the conservatory holds a tropical rainforest and 1,200 free-flying Lepidoptera, Colley says, as a palm-sized blue mountain swallowtail flashes cyan over her shoulder.

The pavilion is seeking a PhD entomologist to become curator, a job that Colley says includes a little bit of everything from research to paying the bills. Colley answers lots of questions from visitors and the public — including how to cure a family's ailing pet scorpion. Pavilion scientists also perform lifespan and breeding studies that will help exhibitions run more effectively. Colley is also a natural with children, which is essential as there are 36,000 school visitors each year.

"You have to know your stuff when kids ask the really big questions," says Olympia Brown, Science for Schools coordinator at the Royal Institution in London. Brown's passion for the history and philosophy of science led her to work in public outreach. She uses her broad view of science to help scientists and educators figure out exciting ways to demonstrate and present ideas to children aged from 8 to 18.

Brown warns, though, that one key skill cannot be learned. "You need the ability to infect other people with your enthusiasm," she says. Both she and Colley suggest testing the water by working as a volunteer at outreach events or museums before seeking a job.

Astronomer Ka Chun Yu, curator of space science at the Denver Museum of Nature and Science, spends his time under the stars in the Gates planetarium and the Space Odyssey exhibition, which combines a lesson on the planets with the flashing lasers of a *Star Trek* set. Yu was originally hired to revamp the planetarium with an interactive, immersive show in which a visitor 'navigates' through a digital simulation of the Universe.

It was a "continuously interesting project", he says, that used both his astronomy background and computer-programming skills. Museum research differs from his graduate work on star formation, says Yu. "I'm blazing a path without much guidance. But I'm doing something that is radically different from what has been done before — and that's important to all scientists."

Kendall Powell is a freelance science writer based in Broomfield, Colorado.

LUCASARTS

J. MORGAN

MOVERS

Philip Bucksbaum, director, Stanford Ultrafast Science Center, Stanford, California



1990–2005: Professor of physics, University of Michigan, Ann Arbor, Michigan

1982–90: Physicist, AT&T Bell Laboratories, Murray Hill, New Jersey

1980–82: Postdoc, AT&T Bell Laboratories, Murray Hill, New Jersey, and Lawrence Berkeley National Laboratory, Berkeley, California

Fateful friendships may have played an important role in Philip Bucksbaum's career, but staying true to his interests in physics has been the key to his success. It was tagging along to a room-mate's physics class, during his first year as an undergraduate at Harvard, that triggered his interest in physics. The class, taught by Nobel laureate Edward Purcell, prompted Bucksbaum's most pivotal career decision: to become a physicist. "That class changed my whole view of things," he says.

Bucksbaum did his physics PhD at the University of California, Berkeley, where a lifelong friendship with Steven Chu — then a fellow graduate student, now director of the Lawrence Berkeley National Laboratory and Nobel prizewinner in physics — would influence his career. After realizing their common interests, Bucksbaum and Chu worked with others in Eugene Commins's lab to test the subtle effects of the electroweak forces at work in atoms.

Graduate research proved a tremendous experience for him. "In grad school somebody else finds the money and you get to devote all of your energy to your experiment," he says. There are few other opportunities to concentrate so single-mindedly on a high-risk, high-payoff project, he adds.

After receiving an offer for a postdoctoral fellowship at the University of Oxford, UK, Bucksbaum gave a talk at Bell Labs in Murray Hill, New Jersey, where Chu had ended up. Bell Labs offered Bucksbaum a postdoc to pursue his interests in picosecond spectroscopy — he turned down Oxford and spent the next decade at Bell. During that time laser pulses became more than 1,000 times faster, drastically increasing the range of fast-moving physical phenomena that scientists can measure.

In 1990, Bucksbaum joined the University of Michigan, where he eventually became the director of the National Science Foundation's ultrafast science centre. He now directs the new Stanford Ultrafast Science Center, where one of the world's first X-ray free-electron lasers is now under construction. He says this technology will not only lead to a revolution in the ability to image small molecules and their motion, but will also have an impact outside physics in disciplines such as chemistry, biology and materials science.

Although key friendships have influenced Bucksbaum's career, he cautions young scientists to do the kind of science they want to do, not something they think they should do. "Have confidence in your own curiosity and be fearless about pursuing it," he says.

Virginia Gewin

SCIENTISTS & SOCIETIES

Back to high school

Researchers spend a lot of time teaching students in universities, but how many venture into secondary schools? Thanks to the Researchers in Residence programme, more than 3,500 PhD students and postdoctoral researchers in Britain have been placed in more than 2,000 secondary schools across the country since the programme began in 1995. They have worked with some 400,000 pupils between the ages of 11 and 18.

Funded by Research Councils UK and the Wellcome Trust, the project aims to inspire secondary-school students by showing them how science can have an impact on people's lives.

It has benefits for researchers as well, such as improving their communication skills. Although many past participants have remained in academic research, they now try to include more public outreach activities in their work. Others can be found in the world of science communication including publishing, the media, hands-on activity centres and museums. Some have gone on to train as school teachers.

Once recruited, researchers discuss with the host teachers how they can contribute to the classroom, making each experience unique. But certain activities have proved popular over the years, such as giving presentations about their research and the latest

developments in their particular field, helping out with laboratory exercises and hosting visits to their research facilities.

Although the researchers are not acting as teachers, they do usually end up doing some teaching. But their decision to teach a lesson is entirely their own. Many start by giving talks about their research, often accompanied by a demonstration and/or an activity for the students to carry out. Often, this leads to teaching a lesson. Researchers typically spend about a day and a half per week doing these activities for an agreed period of time, usually an academic year but sometimes longer.

Since the publication in 2002 of a major review of the British science, engineering and technology workforce, researchers face new requirements for skills training, including communication skills. And with growing pressure from UK research councils on scientists to engage in more public outreach, the Researchers in Residence programme is a good way for young scientists to get started on this skill development. For more information, e-mail: m.m.brodie@shu.ac.uk.

Marilyn Brodie is the Researchers in Residence project manager at the Centre for Science Education, Sheffield Hallam University, UK.

ALUMNUS JOURNAL

The winding road

The path to success in academic science is often fraught with unexpected turns. No matter how much you plan your route, be prepared to make adjustments. During 2004, while I was a Graduate Journal writer for *Naturejobs*, I finished my PhD at Rockefeller University in New York. I decided to move to France for my first postdoc, to gain an international perspective.

I was having a good experience at the Pasteur Institute when an unexpected family situation forced me to return to the United States. Although I enjoyed the laid-back atmosphere (and the cafeteria) at the Pasteur, I learnt that with shrinking funds and a scarcity of tenured or tenure-track academic positions, it is equally difficult in the United States and Europe for new PhDs to penetrate the hierarchy as independent researchers.

I recently began a postdoc at the National Institutes of Health (NIH) in Bethesda, Maryland. Growing up nearby, I often visited the lab at the National Cancer Institute where my grandmother, Alfreda Simmons, worked. She was one of the few African-American women doing lab work at the NIH at the time, and it is fitting that I'm now at the same place, developing vaccines for cancer and HIV.

Travelling down my winding academic career path has taught me how to cope with unforeseen detours. But my goal remains the same: distinguishing myself as a top-notch independent investigator in the not-too-distant future.

Tshaka Cunningham is a postdoctoral fellow at the National Cancer Institute.

For he on honeydew hath fed...

...and drunk the milk of Paradise.

Paul Smaglik

A big, bloated castle on the horizon, bleeding into the setting sun. Hearst Manor. San Simeon.

"I wish we had time to go," I said. In Xanadu did Kubla Kane...

"Oh, Alfred," she said. "You're obsessed."

"I can't help it."

"I know, I know. The same birthdays. The same professions. Except you're not a multi-billionaire. You don't control a vast media empire. You don't start wars to sell more newspapers. Brilliant young filmmakers don't create groundbreaking cinema based on your life. And you don't have sordid affairs that threaten your political aspirations..."

"Rosebud," I said, kissing her on the nose.

"You'd better not, anyway," she said.

"And don't forget to watch the road."

In fact, I could hardly take my eyes off it. The castle was an eye-magnet. More Transylvania than California. More Dracula than Disneyland. Although it did have that Anaheim aesthetic, plunked down in the middle of the Big Sur. The landscape of America. The landscape of nowhere. A history of architecture in one building. NeoGothic PostColonial Revisionism. The art nouveau of wretched excess. Pediments and gables. Towers and balconies. Dormers, cornices, columns. Spires and gargoyles. Things that go bump in the night. The world's biggest and most expensive crypt improbably placed far from graveyards or amusement parks.

Vampire Kane.

"What if he's still alive?" I asked.

"You're crazy."

"No, really — what if he's not allowed to die, or refused to go gently into that good night, as they say, until he's seen everything he's collected, played with all his toys, read all the books, seen all the films, tried on every single suit, laced up and slipped on every pair of shoes?"

She sighed and shook her head.

"It's quite possible," I insisted. "The miracles of modern medicine and all."

She looked away suddenly and stared intently out of the window.

The wrong thing to say after reproductive technology recently let us down. But it was true, theoretically. Cellular immortality, thanks to telomerase. Keeps the DNA from getting frayed, like the ends of an old rope. So the cells can divide over and over and over again, with no defects. A genetic photocopier with amazing resolution.



Not yet available at a hospital near you. Not yet. But they already sell spare parts now, hearts and lungs, kidneys and livers, grown up from embryos, fetal tissue, skin scraped off the tip of your own nose. They have waiting lists, but it's amazing what a little cash can do. Bump you up to the front of the line. Strictly hush-hush money, of course. Especially if you're already supposed to be dead, not just going through desperate measures to fend it off due to vanity and insecurity. Although that probably has something to do with it. Along with the ability to do it.

Why climb Mount Everest? Because it is there.

Why live forever? Because you can.

A tune-up every 100,000 miles or every ten years, whatever comes first. The money's in the bank, the cheque's in the post. Synthetic joints, artificial cheek bones, stainless-steel supports, rack-and-pinion steering. Gives 'a new lease of life' a whole new meaning.

Just call in the docs whenever necessary and nip and cut and bait and switch and presto! A whole new circulatory system! Nanotube technology, brimming with synthetic blood. Or maybe a new skin. After a few months of healing (what are a few months when you've got forever?) unwrap the gauze and remove the i.v. drip and take a look around the pharaoh's tomb, somewhere deep within the bowels.

Then throw the bandages aside and start taking stock.

I can picture him, appearing from some nook after the last tourist has left, slipping the guard a Benjamin from his unlimited supply. He'd have to read the newspaper first. A Hearst publication, of course. Then the competition's. Then all his press clippings of the day. Everything germane to the company. Which is pretty much everything. But that's the way it goes when you've got a multinational, horizontal conglomerate. Everything connects to everything else, somehow. Everything is relevant.

By the time he has processed all the day's news, he will sense the sun rising, take a quick peek at some crated work of art, some great hidden Picasso, maybe, sigh, then disappear for the day. He'd never be done. He'd barely ever even get started. Maybe he still controls his empire. By proxy. By secret decree. Maybe there's a series of Venn diagrams in a safe somewhere covering every possible decision. Or flow charts. Marching orders. Battle plans. It can't be that hard, anyway. Buy. Build. Expand. Conquer. People do it every day. They have their names on buildings to prove it.

Paul Smaglik edits the *Naturejobs* and *Authors* sections of *Nature*. This excerpt is from an as-yet-unpublished novel with the working title of *Monument*.

Abstractions



FIRST AUTHOR

Teasing apart the various effects of global warming is tricky because there are so many variables to consider. This is especially true when assessing the effects on

glaciers and ice caps. Some scientists say that global warming is making these bodies melt faster, contributing to rising sea levels. But many of these assessments have used relatively simple models of glacier and ice-cap size and so may not have captured the complexity of the dynamics involved.

Page 311 features the results of a more sophisticated model built by Sarah Raper of the Centre for Air Transport and the Environment, Manchester Metropolitan University, UK, and Roger Braithwaite at the University of Manchester. This model suggests that the global rise in sea levels caused by global warming might not be as great as some estimates have found. But it also shows that local effects, especially in high mountain areas such as Nepal, could potentially be catastrophic. Raper spoke to *Nature* about the results.

Were you surprised that the rise in sea level predicted by your model was up to half that predicted by others?

I didn't have previous expectations. I like this paper because I think it gives a much better framework for glacier and ice-cap melt. It is ongoing science and we are still filling in lots of gaps. We have to extrapolate to the globe what we know from only a few glaciers.

Your paper only models one aspect of sea-level change. What are some others?

We think that the other most important aspect is the thermal expansion of the ocean, in which the ocean warms up, expanding its volume and so pushing up sea level. We're not modelling that. Two other possible factors are the large ice sheets in Antarctica and Greenland, which could contribute a great deal, but it's questionable how much of an effect they would have over the next 100 years. This is actually a big unknown.

Is there any danger that your conclusions might be leapt upon and used by someone with a particular political agenda?

I am not at all belittling the importance of global warming. It's just that our estimate of one particular aspect is a bit lower.

How did you divide the work with Roger?

I approached this from a climatological point of view. Roger is more of a glaciologist; he's got much more field experience.

What's next?

We're looking for more observational data that can be easily assimilated into our model. We should look at the possible effects of changes in precipitation. ■

MAKING THE PAPER

Nicholas Katsanis

Genetic analysis of an inherited disease gives voice to a polypeptide.

Ever since he did his postdoc, geneticist Nicholas Katsanis has been fascinated by the way a given genetic disease can affect people in many different ways. Now based at Johns Hopkins University in Baltimore, Maryland, Katsanis wanted to know how a single disease can have so many symptoms and outcomes in patients, even when those patients are related. On page 326 of this issue, Katsanis and his colleagues present some clues as to what may be happening.

Katsanis chose to investigate a genetic condition called Bardet–Biedl syndrome, which can cause symptoms such as obesity, vision problems and cognitive defects. The disease shows quite dramatic variability — even within members of the same family, its severity can range from mild to profound. Moreover, it affects organs and systems within the body that have no obvious links — for example, there can be both hormonal defects and extra digits.

Originally, it was thought that the syndrome was linked to just one gene. But when the human genome sequence was completed, several genes were found to be involved — and in some cases mutations in two genes 'cooperated' to cause the disease and modulate its severity. Despite this advance, researchers were still unable to explain all the variations between patients.

Geneticists are now fairly good at starting with differences between people and working out the links to a particular type of disease — in other words, going from phenotype to genotype, says Katsanis. But to understand variability within a single disease means you have to work in the opposite direction. "We can go from phenotype to genotype fairly well," says Katsanis. "Where we are significantly challenged is going from genotype to phenotype."

The work by Katsanis and his team on



Bardet–Biedl syndrome illustrates why this approach is so difficult. They had to negotiate four separate threads of research, each of which presented them with challenges. Starting with the premise that all the proteins involved with the disease participate in the same cellular process, they identified numerous proteins that might be involved. Further analysis narrowed the candidates down to one polypeptide, and its involvement was then confirmed in zebrafish.

Even though Katsanis and his team had identified the culprit, they still needed to work out how this short sequence of 241 amino acids could cause such wide variability among sufferers of Bardet–Biedl syndrome. They sequenced the gene and undertook an intensive round of functional experiments. This allowed them to see that a single, mild variation in the gene interacts with disease-associated mutations in other genes, which cumulatively can radically change the disease's severity and symptoms.

Although he is pleased with the result, Katsanis notes that it leaves us with a sobering thought. The gene that his team identified as key to the disease's variability was thought to be 'silent' — in other words it wasn't expected to be responsible for any effects. "The scary thing is, how many sites like this exist in the human genome?" Katsanis asks. "I think there are going to be thousands — which means we no longer have the luxury of readily discarding these silent changes." ■

QUANTIFIED SWEDEN

A numerical perspective on *Nature* authors.

Uppsala University in Sweden has proved to be the ideal choice for Martin Brazeau's PhD on vertebrate evolution. The department offers a world-class research environment, headed by Brazeau's supervisor, Per Ahlberg, and the atmosphere is lively and interactive, Brazeau says.

Although everyone has their own project, there is extensive collaboration. "Our projects are close enough that we act as good sounding-boards for each other's ideas," Brazeau explains. Brazeau's own studies focus on the Palaeozoic era (540 million to 250 million years ago), particularly early bony fish and tetrapod evolution. On page 318, he describes some of his latest work with Ahlberg, which looks at the earliest stages in the evolution of the middle ear of terrestrial vertebrates.

7 authors working in Sweden report findings in *Nature* this week.

4 original research submissions to *Nature* made during the first two weeks of January 2006 came from Sweden (total number of submissions = 198).

5 other PhD students and postdocs are working with Brazeau in Ahlberg's research group at Uppsala.

1477 is the year Uppsala University was founded, making it the oldest university in the Nordic countries.